

10 December 1981

Not a revolution, more a Trojan horse

The House of Lords committee on science and technology has produced a guileful recipe for change. Will the British government listen to what it has to say?

In the nineteenth century, when the British government felt threatened from abroad, the natural response was to build a few more battleships. Now, when most threats are economic rather than strategic, the temptation is to carry through yet another reorganization of the administration of science and technology. Untrue to form, the present British government has been restrained almost to the point of seeming indifferent on the issue, and masterful inaction may indeed be appropriate. For although there is a long-term connection between national prosperity and the efficiency with which research and development is administered, the short-term consequences of tinkering are usually disadvantageous. Yet it is also plain that the machinery of research administration in Britain has for years been out of date. The British government should thus be grateful for the thoroughly sensible report of the House of Lords Committee on Science and Technology, published this Monday. Whether gratitude will flow is naturally quite a different matter. Indeed, from the point of view of a government so embattled on other fronts that it may be forgiven for seeking respite where it can find it, the advice it has received from the House of Lords has the unwelcome attributes of being at once persuasive and practicable.

More clearly than its predecessors, the House of Lords committee has distinguished between the two functions of science (and scientists) in government — the management of research and development and the provision of scientific advice on broader questions of policy. The distinction has been made before, most recently by Dr Martin Holdgate's inquiry into the role of the scientific civil service within the British government, but it has not sunk in. The point is, however, both simple and crucial. Government departments sustain (or should sustain) substantial programmes of research and development and should manage them efficiently. They also need some mechanism for being told from within which policies make scientific sense and which do not. Lorth Rothschild's proposals in 1972 for the reorganization of civil science supposed that if the spending departments of government were provided with chief scientists of sufficient stature, both functions could be carried out by the same people. The experience of the past decade has, however, shown that most chief scientists have been imprisoned by the management of research and development. The House of Lords suggests that this danger could be avoided by appointing two people within each ministry, one to advise on general policy and one to manage the research programme. That is one solution to an immediate problem. A better one would be to ensure that chief scientists should be strengthened, and equipped, to carry out both functions. Either way, something must be done, both departmentally and for central government.

The way in which chief scientists in British government departments have been converted by the pressure of work from putatively subversive agents of change into high-grade clerks is not, however, the only failure of the post-Rothschild era. Even at the outset, it was clear that the Rothschild reorganization entailed the risk of emasculating crucial parts of the previous machinery for providing the government with scientific advice. Thus the old Council for Scientific Policy, which had become fond of writing

(and, to its credit, publishing) reports on matters of general importance, was replaced by the Advisory Board on the Research Councils, a referee to arbitrate between the research councils which appears even to have given up the practice of saying how it purports to do so. The nearest entity within the British government to a central source of advice on science policy is the committee known as the Advisory Council on Applied Research and Development — a busy committee conspicuous chiefly for its lack of power and coherence. The House of Lords committee proposes that this committee should be transformed by some kind of evolutionary process into a Council on Science and Technology, that it should report to some member of the British Cabinet who, while not being a Minister of Science and Technology, would be spokesman in this cause and that the committee should be serviced by the member of the Cabinet Office with responsibilities in this field while its chairman should have access to the nominated minister.

The proposal is sensible. In British terms, it is also devilishly clever. Strengthening a committee is, after all, a good thing to do; setting up a new committee would be unpalatable. Giving an existing minister wider responsibility will seem productive; appointing another would be counted as an encouragement of the growth of bureaucracy. The House of Lords committee has also been politic in the modesty of its requirements of defence research and development. It asks not that there should be a means by which the true ivory tower of British science (defence research) should be coordinated with the rest, but merely that there should be cross-representation on the appropriate committees and that everybody should bend his energies to ensure that what the government spends on defence research should not continue to have no commensurate civil benefit. Cleverly, however, the committee asks that its new council should have some right to ask questions about the interface between defence research and civil industry. It would, of course, have been impolitic to recommend that the committee should have the right to ask what defence research is for.

With a little luck — and if the committee's friends do not too eagerly disclose their reasons for believing that it is right — the outcome may be a more intelligent way of running that part of British science and technology supported by the public purse. That is as it should be. Pinstriped gentlemen have for too long been turning up at pointless committee meetings in order to agree that it would be better if there were time to talk about matters of substance, not of procedure or of who gets how much. That the committee's recipe for the administration of civil science is to return to the 1950s, when there was a cabinet minister with part-time responsibility for science and technology and an advisory council whose power derived from its reputation for plain speaking (but not good judgement) is not in itself a reason why the recipe should not be tried. For something, even anything, must be done. It cannot be in the private interests of even British governments that they should feel free to decide, without internal dissent, to rob the academic research base of essential support, discovering only afterwards that they have erred. That, however, is what is now in prospect.



US dispute on nuclear bomb safeguards

Agency and Administration at loggerheads

Washington

The State Department and the Nuclear Regulatory Commission (NRC) are drifting towards loggerheads on anti-proliferation safeguards. Both parties agree that the safeguards administered by the Vienna-based International Atomic Energy Agency (IAEA) need to be improved. The State Department, reflecting a general drift away from the policies of the Carter Administration, would prefer to use conventional diplomatic channels.

In contrast NRC, which under the Nuclear Non-Proliferation Act signed by President Carter in 1978 is required to assure Congress about the adequacy of IAEA safeguards, is pushing for a more direct approach. Following sharp criticism of IAEA procedures by two former inspectors (*Nature* 26 November, p.300), the five members of the commission have sent a letter to all members of Congress saying that the current safeguards system "would not detect a diversion in at least some types of facilities".

In the light of this concern and the general shadows that have recently been cast over the adequacy of IAEA safeguards, NRC has decided to carry out a review of its own policies on granting export licences for nuclear materials. The State Department, however, although embarrassed by disclosures about weaknesses in IAEA's safeguards system, does not seem to feel any such drastic action is necessary. Addressing the Senate Foreign Relations Committee last week, Mr Richard T. Kennedy, Under-Secretary of State for Management and a former member of NRC, repeated the Administration's commitment strongly to support the agency.

Introducing the Senate hearings, committee chairman Senator Charles Percy emphasized the central importance of IAEA, and pushed Mr Kennedy hard on whether the Administration was doing everything it could to prevent the spread of nuclear military technology, complaining that the Senate had been provided with insufficient information on procedures agreed with Pakistan and IAEA to prevent the diversion of civilian technology. Senator Percy put the suggestion — strongly denied by Mr Kennedy — that the Administration might be going "soft" on non-proliferation to help boost the sagging exports of the domestic nuclear industry.

Democrat Senator John Glenn went

even further, suggesting that the State Department should put pressure on other industrialized countries to provide a significant increase in technical and financial resources for the agency. To support his argument that such increases were badly needed, Mr Glenn produced a report prepared at the Pacific Northwest Laboratory under contract to the Department of Energy throwing doubt on whether IAEA was able to achieve its technical objectives of the surveillance and accounting of nuclear materials.

One issue on which there was little disagreement between Senator Percy and the Administration was over attempts by

Third World countries to gain greater influence on the activities of IAEA, both in terms of increasing their membership of the governing council and in appointing more inspectors from the developing nations. On the latter point it was agreed that, although an important objective, it should not be pursued at the price of sacrificing the quality of inspection procedures. On the other hand, Mr Kennedy made it clear that the State Department would resist attempts to "politicize" the agency, in terms of changing the current balance of control.

There is a deeper division over the extent to which a lack of detailed information about the technical data used by IAEA in

Is creation a science or a religion?

Little Rock

The state of Arkansas was accused on Monday morning of making "an unprecedented attempt to legislate what is science" at the opening of a trial over its efforts to determine how different theories of the origins of man should be taught in its public schools.

The American Civil Liberties Union (ACLU) is challenging as unconstitutional a law, passed by the state legislature earlier this year, which requires that whenever the Darwinian theory of evolution is taught, an equal amount of time should be devoted to presenting evidence in favour of what is described as "creation science".

ACLU contends that the law violates the first amendment to the constitution, which requires a strict separation of church and state. It argues, in the words of Little Rock attorney Robert M. Cearley, that "creation science is not science but religious apologetics, an attempt to prove or justify sectarian religious beliefs". The state argues that there is nothing inherently religious in teaching that the world came into being as the result of a positive act of creation, and that the evidence used to support this hypothesis is "at least as scientific" as the evidence for the "evolution science".

Setting the agenda for a debate that is likely to dominate the trial, state attorney-general Stephen Clark responded in his opening statement that "this is not a trial about religion but a trial about science".

In 1968, the United States Supreme Court struck down as unconstitutional a state law forbidding the teaching of evolution on the grounds that it contradicted a literal interpretation of the Bible. This time the issue has been reversed. ACLU has brought the case on behalf of 23 local religious leaders, biology teachers and schoolchildren, claiming that the new law infringes their rights to freedom of religion.

"For practical purposes it is a tremendously important case for us because it is the first of its kind, and the

decision of the court will get a lot of attention", said ACLU staff attorney Jack Novick before the case started. He pointed out that a similar bill has recently been passed in Louisiana, and that others are pending in up to twenty more states.

Three complaints are being made by ACLU. The principal charge is that the law violates the first amendment to the constitution. It is also being charged that the law infringes the academic freedom of teachers and schoolchildren, since it imposes conditions under which evolution can be taught; and that the language in the bill is unconstitutionally vague.

The state attorney-general in disputing each of these charges claims that, since there is nothing necessarily religious in the idea that the world came into being at the hands of some undefined form of creator, there is no problem with the constitution.

Each side will introduce a string of scientific witnesses to support its case. For ACLU, these include prominent academics in fields such as biology, geology and palaeontology; the state plans to introduce qualified scientists who claim their results substantiate the creationist theory of origins. Furthermore, unlike previous cases in which the relative merits of the two explanations were directly compared, this time the focus will be as much on the philosophy of science, with Judge William Overton deciding what types of ideas and reasoning can be described as "scientific" and "religious".

The case is expected to last for two weeks. If ACLU loses, it is already promising that it considers the case so important that it will take it as far as the Supreme Court, which has ultimate say on constitutional issues.

If the civil liberties group wins, then Mr Novik says it will press for full payment of its legal costs — calculated to be several hundred thousand dollars — against the state of Arkansas. But in that case, the state is expected to appeal against the verdict. So either way, the full legal debate is likely to last for a long time.

David Dickson

reaching many of its conclusions makes it difficult to check on whether the agency is doing a good job. NRC member Mr Victor Gilinsky expressed concern that protection given in IAEA statutes to industrial, secret or other confidential information has been interpreted in the broadest manner to withhold information about the safeguards system.

Suggesting that undue secrecy made it difficult for NRC to fulfil its responsibilities for checking the adequacy of IAEA safeguards — responsibilities whose legal status still remains ambiguous — Mr Gilinsky said that the commission believed more information should be reported to IAEA's board of governors by its secretariat, and that the United States should try to obtain better information on overall international safeguards effectiveness and on any significant safeguards problems in countries receiving US exports.

Mr Kennedy replied that the general limitations on the agency's role — such as the substantial amount of classified information which it could not release — are "simply facts of international relations. . . The IAEA safeguards system entails a unique compromise of sovereign rights by many nations, and it is certainly no surprise that this compromise is subject to specific limitations". **David Dickson**

Germany's breeder Gap unbridged

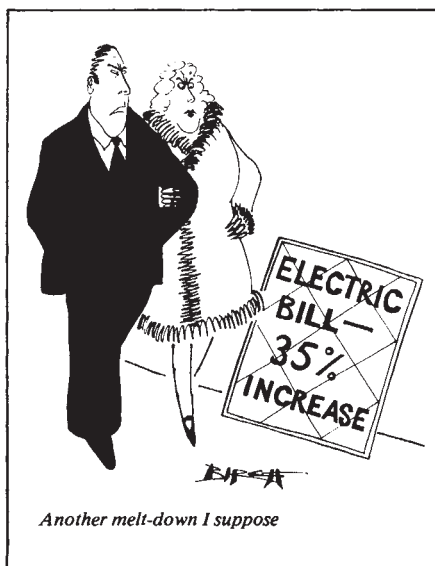
Ulm, West Germany

Rapidly rising costs are threatening the future of West Germany's SNR-300 fast-breeder reactor, with government and public service utility companies still arguing over who should foot the increased bill if the project is to continue.

The Minister of Research and Technology, Andreas von Bülow, has succeeded in extracting from the utilities an increased financial contribution to the SNR-300 plant at Kalkar. At the end of October he announced that Rheinisch-Westfälisches Elektrizitätswerk (RWE), the leader of the owner/operator utility group, is willing to put up an additional DM 375 million. Two other utilities not so far involved in the project, Preussenelektra and Nordwestdeutsche Kraftwerke (NWK) are prepared to contribute DM 172 million.

Even so these pledges fill only half of a total shortfall of DM 1,100 million created by recent cost overruns. The federal government is unwilling to provide the extra funds from its own budget, and insists that the full amount should be paid by West German industry and in particular by the utilities, each financing a fraction corresponding to its share of electricity generation. "It is perfectly clear," von Bülow recently said, "that without an increase in the participation of the utilities, the future of the fast breeder is in danger and a halt to construction before the end of the year cannot be ruled out".

The weeks ahead will show whether von Bülow's warning is merely a tactical ploy or whether there is a real chance that work on the fast breeder will stop. Although budgetary problems have triggered off this latest confrontation, if the minister regarded the fast breeder as an absolute economic necessity it seems likely that he could accommodate the necessary additional DM 550 million over five years (DM 110 million a year) within this present budget — the ministry has a proposed total budget of DM 6,600 million for 1982 and an energy research budget of DM 2,550 million. "The present situation," he said, "necessitates a reconsideration of the original objectives of government support." At the same time the minister announced that government funds for microelectronics and biotechnology will be increased over the next few years.



Political uncertainty is one of the reasons why the utilities are reluctant to shoulder an increased share of SNR-300 costs. The Bundestag has reserved to itself the decision on whether the SNR-300 should be put into operation when completed. The utilities hesitate to commit themselves as long as this uncertainty persists, and the pledges made so far are contingent, among other conditions, on a positive decision from the Bundestag.

Licensing uncertainties have been another major obstacle — but here at least the greatest difficulties appear to have been overcome with the passing of the fourth part of the licence on 15 October.

Some utility companies originally argued that SNR-300 is a national project that should be financed from taxpayers' money. It remains to be seen whether they will soften their stand in the face of a threat to abandon the project. The West German utility association, VDEW, is now discussing alternative finance strategies including a research and development levy on electricity sales modelled on the present subsidy for West German coal used in the generation of electricity. **Otto Keck**

Israeli education More gloom

Tel Aviv

The complaint that cuts in higher education could seriously endanger Israel's scientific potential in the 1990s ran through the symposium organized at the end of November by the Israeli popular science magazine *Mada* to celebrate its twenty-fifth anniversary. Ironically, the symposium coincided with the *Isratech* exhibition, a celebration of recent Israeli achievements in applied science.

While *Isratech* celebrated the success of Israel's funding policy, which for the past decade has emphasized applications, symposium participants maintained that, unless investment in higher education is stepped up, Israel's advantage in many sectors of technology could prove short-lived. The symposium was chaired by the physical chemist Dr Ephraim Katzir, formerly Katchalski, who produced the 1968 report which effectively switched government emphasis from pure to applied research.

Unlike other new states, Israel has a research infrastructure which antedates independence and which is based on institutions such as the Technion, the Weizmann Institute, the Hebrew University and the Volcani Agricultural Institute. Until the late 1960s, Israeli science policy was aimed at more of the same. The Kachalski report redirected government attention to the needs of industry. Only some of Katchalski's recommendations were implemented. Chief scientists were established in ministries involved in research and development, but the proposed upgrading of the National Council for Research and Development failed to materialize — indeed, in 1977, it was actually downgraded, being transferred from direct responsibility to the Prime Minister's Office to the new Ministry of Energy and Infrastructure. At the same time, all spending on research has been eroded by inflation, now 133 per cent a year.

Even so, the post-Katchalski era has produced a number of achievements. Industry has done well, and now employs more science graduates. Agriculture has seen a steady stream of advances in crop breeding, soil chemistry, irrigation techniques and such novel methods as the sterilization of soils by solar energy. Both production and productivity have increased.

The symposium was chiefly alarmed by the other side of this coin — the continuing absence of a national science policy and the continuing brain-drain to the United States and Europe. The participants blamed people's willingness to leave on the cutbacks in university finance and the mismatch between what universities teach and what industry (and students) demands.

In the past ten years, there has also been a stagnation of student numbers in the exact sciences while total student numbers have soared. The underlying reason is not

clear, although the decline has been greatest in the "traditional" fields such as civil engineering, while the most modern departments such as computer sciences are full to overflowing. This suggests that the universities are failing to adapt to modern trends — an ironic conclusion when they themselves have founded such institutions as the Institute of Forward Planning at Tel Aviv or the Neimann Institute at the Technion to act as potential policy advisers.

Employment problems are exacerbated by the relatively small payrolls of the new science-based industries. Would-be students therefore think twice before investing some five or seven years — delayed by a further three years military service — in an exact science when the job prospects are so uncertain.

The Jerusalem College of Technology, which was specifically founded to train the expert workforce the country needs, may be an exception. It is a relatively small and religious foundation where the all-male students study Judaism in the mornings and technical subjects to first degree level in the afternoons and evenings. The "new pioneering spirit" described by the founder and rector, Professor William Low, has presumably helped to take the edge off job uncertainty.

Perhaps inevitably, the symposium produced a gloomy forecast for the next couple of decades. According to Professor Joshua Yavner of Tel Aviv University, unless the government stops the cuts in higher education, Israel will suffer the loss of its main "natural resource" — educated manpower. Hitherto, the University Grants Committee, which administers the government's block funding of higher education, has tried to restrict the cuts to ancillary and non-academic fields (although library budgets have been considerably curtailed), but a point will soon be reached where no more such economies are possible. Then, inevitably, the output of science graduates and PhDs will begin to decline, and the new science-based industries, founded on the expertise of those who graduated in the 1960s and early 1970s, will fail to achieve their early promise.

Vera Rich

US research patents

Conflicts arise

Washington

United States universities are lobbying hard in defence of their newly-won right to an automatic patent on all inventions arising from federally supported research. They are alarmed that their rights may be circumscribed in what may be the most profitable field — genetic engineering.

Two weeks ago, the House of Representatives Science and Technology Committee approved an amendment to a draft bill on patent rights, already approved by the Senate, to exempt the results of research using recombinant DNA techniques from

the broad provisions of patent legislation passed by Congress last December.

The amendment was offered by Representative Albert Gore of Tennessee, chairman of the committee's investigations and oversight subcommittee, and follows his criticism of an agreement between Massachusetts General Hospital (MGH) and the German chemical company Hoechst that would, he claims, allow the company to benefit unfairly from research paid for by the US taxpayer (*Nature* 18 June, p.525).

Several Washington-based groups representing research universities have protested, saying that this amendment undermines the whole purpose of the new legislation, which is to speed the transfer of research results to private companies. They point out that three years ago, after consulting universities and industry, Dr Donald Fredrickson, then director of the National Institutes of Health (NIH), concluded that no special arrangements need be made for the application of patent laws to genetic engineering research.

An aide to Mr Gore said last week that the congressman might agree to withdraw the amendment before the draft bill reaches the floor of the House for discussion. The Senate has already passed such a bill which even expands the scope of last year's legislation to include not only universities and small businesses but all companies receiving federal research funds, and Senate aides said last week that they were totally opposed to the House committee's amendments.

However, Mr Gore still seems keen that the bill should make a sharp distinction between the control of research results funded from public and private funds. This could complicate efforts by NIH to interpret the new law's provisions on jointly funded projects.

Under the terms of the MGH/Hoechst agreement, the company will provide the hospital with up to \$50 million over a ten-year period to support basic research carried out in a new Department of Molecular Biology. All the research would be under the control of MGH and all research results could be freely published after a 30-day preview by the company, but Hoechst would have first refusal to license patentable results.

Both MGH and Hoechst insist that this arrangement is compatible with existing academic practice, that they have gone to great lengths to avoid any possible conflicts of interest or undue secrecy, and indeed that the terms of the agreement might act as a model for other universities seeking research support from the private sector.

Mr Gore, however, says that his concerns are endorsed by a report he commissioned from the General Accounting Office, the investigative arm of Congress, which pointed out ways in which the agreement might conflict with the patent legislation passed last year. Under the current legislation, any patents arising from jointly

funded research would remain the property of the research institution, but the federal government would be permitted free use of the patent, and would also be able to reclaim control of the patent rights. However, MGH acting director general Dr Joseph Martin insisted that a clear distinction would be made between research supported by the federal government and that financed by Hoechst.

Mr Gore remains unconvinced. He argues that maintaining a clear separation between separately-funded research projects is "simply a fiction" and promises that, even if his amendment to the patent legislation is dropped, he will be keeping a close eye on the MGH/Hoechst and other similar agreements, to protect what he describes as a massive public investment in recombinant DNA research. An official in the legislative office of NIH said last Friday that he felt the concern raised by the General Accounting Office; about jointly funded research "is certainly a problem that is not non-existent, although it does not currently appear very prevalent". A delegation from NIH will be visiting MGH next week to discuss the impact of the new patent legislation on future arrangements for federally funded research at the hospital.

David Dickson

Science in France

Democratic union

Grenoble

How far can democracy go? This question is emerging as one of those central to the new politics of science in France, and it was put to a remarkable test last week.

Professor Robert Tournier, the 47-year-old director of the renowned laboratory for very low temperatures at Grenoble, in the Rhône-Alpes, announced some time ago to his colleagues that he wanted to resign the directorship to have more time for research in his own laboratory. Normally, the CNRS would take advice on a new director from the laboratory steering committee, and from the CNRS "parliament", the Comité National, and then would merely announce an appointment. This time it has another factor to take into account. Internal advisory committees in the laboratory could not agree on their recommendation for a successor, and Fournier, as a good democrat and socialist, asked the whole laboratory, including technicians, to give their advice on two candidates.

The resulting "election" took place last week. It has no official significance, and was nearly a draw. But the trades unions (their favoured candidate having won) are treating the event as a clear selection of director, and have written to the Paris headquarters of the Centre National de la Recherche Scientifique, to which the laboratory belongs, to say so.

As it happens, the narrowly-elected

candidate — a personable 39-year-old physicist called Daniel Thoulouze — is also likely to be the recommendation of the relevant section of the Comité National, so there may be no conflict. But the CNRS directorate is put in an awkward position whatever the result. If it selects Thoulouze, it will appear to approve of the grass-roots election of laboratory directors — which it wants to avoid. Yet if it selects another, it will appear to confirm the fears of the more radical scientists and the strong trades union movement in French science that, despite a socialist government, the selection of laboratory directors — who have great power over their laboratories — will remain closed and centralized.

In the event, the CNRS may be forced to ask Tournier to continue. The new CNRS director-general, Jean-Jacques Payan, is not against elections in principle; but he says they should be undertaken only "in a very precise framework", meaning that the electorate should be appropriate to the category of post. But Payan is under great pressure from the grass roots, the technicians and the younger researchers, for a more democratic way of appointing directors, and he will certainly have to give in to some extent. As Tournier himself said this week, the question is whether the CNRS is forced into a system where directors must conform to majority opinion in a laboratory, or a system where the directors can be independent, but must subsequently seek to bring the whole laboratory behind them.

Meanwhile the unions of Grenoble, which have enjoyed a large degree of democracy (or at least, consultation), since 1968, are determined to be "serious" about the election. The candidates were grilled in debates, and their scientific and general laboratory policy examined, they said. The unions' objective is that the laboratory should continue to do good science, and they insist that that requires the director to have good relations with his staff, down to the lowest technician.

And to some extent, this policy is reflected at the Ministry of Research and Technology in Paris, which gave Payan his job at the CNRS. The strong feeling there is that science is more and more a collective exercise — and that democracy and efficiency go together. But on the other hand, they also feel at the ministry that too much democracy leads to "turbulence", a position which would seem to be close to Payan's own (which is perhaps not surprising, for the minister for science, Jean-Pierre Chevènement, and Payan have been close colleagues for a decade.)

In the end, the issue is going to be how far Chevènement is prepared to travel with the unions — and particularly the communist-affiliate CGT, which is strong at Grenoble. The CGT secretary here said last week that he and his union were watching and waiting. And if Chevènement does not come up to their expectations, he promises a struggle.

Robert Walgate

What price IIASA?

Washington

The Vienna-based International Institute for Applied Systems Analysis (IIASA) has now been informed officially that the United States will be withdrawing its support for the institute from the beginning of 1983 (see *Nature* 3 December, p.390). This gives IIASA some breathing space in which to seek ways of keeping the United States involved — even if its contribution is severely reduced.

At present the United States and the Soviet Union together provide about half of IIASA's \$10 million annual budget, and the IIASA council is now looking at ways to reassess the dues from each of the member nations. At present the United States's \$2.5 million contribution comes from the National Science Foundation and is channelled through the National Academy of Sciences, and one possibility is that the money might be raised from other federal agencies. Private foundations are also being approached but so far all offers for support have been well below current levels.

Whether the United States will come back remains an open question. Founded in the early 1970s as a symbol of détente, IIASA has provided a useful window for both East and West on economic and social conditions on the other side — for example information about energy resources. But the institute has failed to stem doubts about the true academic value of its work.

David Dickson

Pakistan's space ambitions

A military option?

Bangalore

Pakistan looks set to embark upon an ambitious programme of space research — with clear hints that the spur is the prospect of military applications in the long term. Dr Salim Mahamood, chairman of Pakistan's Space and Upper Atmospheric Research Commission (SUPARCO), while unveiling the details of a 10-year national space programme in Islamabad recently, said that Pakistan did not want to be left behind in the space race, and that his country is studying in detail the configuration of a satellite which "can serve strategic purposes by taking pictures of military installations, army movements and acting as control, command and communication bases".

The setting up of SUPARCO earlier this year, with President Zia at its head, is said to have given fresh impetus to the hitherto slow-moving Pakistani space programme which is currently concerned mainly with launching sounding rockets from a base on the outskirts of Karachi. Dr Mahamood

claims that Pakistan is on the way to developing a communications satellite that "would be launched with the aid of either a European or an American rocket".

Since the successful flight test of India's SLV-3 rocket in July 1980, there have been fears in Pakistan that India might use experience gained with SLV-3 to build a strategic missile. Indeed, in November 1979 an Indian parliamentary committee was told that within six months of a political decision, SLV-3 could be modified into a medium-range ballistic missile. But the Indian Space Department stresses that military objectives are beyond the purview of the Indian space programme which is directed at "harnessing space for socio-economic development".

There is concern too that with the launch of the Bhaskara II satellite, India is developing the ability not only to survey the land for natural resources but also to carry out military surveillance over Pakistan.

Dr Mahamood claims that over the next five years, Pakistan will attempt to test a space launch vehicle of its own. The controversial deal between Pakistan and the West German private enterprise space agency OTRAG seems to be in obedience. It is claimed that Colonel Gaddafi, OTRAG's host in Libya, had contacted Pakistan with an offer of OTRAG's technology.

How true is the rumour of a link between OTRAG and Pakistan is not known but some strategic analysts say that OTRAG could provide about 2,000 qualified space technologists badly needed to bolster Pakistan's space research efforts.

Meanwhile, Pakistan is negotiating with the US National Aeronautics and Space Administration for the setting up of a \$10 million ground station near Islamabad to receive earth resources data from the Landsat remote sensing satellite.

B. Radhakrishna Rao

India's space programme

Now for real

New Delhi

India entered a new phase in its development of a series of satellites for remote-sensing applications with the successful launching of its experimental satellite Bhaskara II, into a circular Earth orbit from the Soviet Union on 20 November. India's space programme for the 1980s includes as one of its two chief goals a major survey of natural resources. The other task is the development of space-based communications system. The first operational satellite, the National Satellite 1A will be launched in April 1982.

The 440-kilogramme Bhaskara II is an improved version of Bhaskara I, which went out of action, after 26 months, in August. Modifications to the camera system should eliminate the "corona discharge" which caused the breakdown of one of the two television cameras aboard Bhaskar I for a few months. And an extra

frequency has been added to its radiometer for more accurate measurements of atmospheric water content, a key factor in the satellite's subsidiary role of weather forecasting. The satellite also carries indigenously developed solar power cells for space qualification tests.

The next step after Bhaskara II will be the 600-kilogramme Indian Remote Sensing Satellite I (IRS I). A memorandum of understanding has been signed with the Soviet Union for launching it in 1986.

But India now has to face the prospect of a much more expensive space programme than in the past, since Bhaskara II marks the end of free launch vehicle and rocket booster facilities provided by the Soviet Union since the launch of the first Indian satellite in 1975.

Indian scientists are working on indigenous launch vehicles as alternatives, although experiments with the SLV-3 launch vehicle have so far achieved only qualified success. Nothing daunted, the Indian Space Research Organization is developing the Augmented Satellite Launch Vehicle (ASLV), based on SLV-3, which can put a 150-kilogramme satellites into space. India also plans to develop a polar satellite launch vehicle by 1986-87 to launch 600-kilogramme satellites into sub-synchronous polar orbit 500 to 100 kilometres from the Earth. **Sunil Saraf**

European environment

One log-jam ends

Brussels

The Council of EEC Environmental Ministers which met in Brussels on 3 December was remarkably successful, with agreement being reached on several previously contentious issues.

The directive regulating discharges into the aquatic environment of mercury from the chlor-alkali industry was adopted. Attempts to push through the mercury directive and the "drins" directive (regulating aldrin, dieldrin and endrin) foundered on the dispute over whether quality objectives (favoured by the British) or limit values (preferred on the continent) should be used to assess the maximum permitted level of pollution discharged into water from an industrial plant. The ministers have now agreed on a clause requiring the latest technology to be used when building new plants.

The haggling over the Seveso directive on the prevention of major industrial accidents also ended at the meeting. The new French government has toned down its objections to the clause on the consultation of countries across the frontier from new plants. The consultations will now take place on a bilateral basis — without the involvement of the European Commission as originally proposed. Although the directive does not cover nuclear installations, the same consultation procedure is at the centre of a dispute between France and

Electronic intrigue

Palo Alto

In probably the largest theft of semiconductor devices ever, \$2.7 million worth of integrated circuits disappeared over Thanksgiving weekend from Monolithic Memories Inc., a "Silicon Valley" company.

About 498,000 chips of two types were stolen. One kind, a programmable array, PAL, embodies circuits that provide a logic function in a wide range of computer products including video games, industrial equipment and computers used by the military. The second type, FIFO, is used to input data at one speed into computers and take it out at another.

About \$20 million worth of advanced computer chips disappear each year into a national and international underworld electronics trade. Integrated circuits are smaller than a fingernail, hard to identify and can sell for \$100 apiece. Many stolen chips are smuggled out of the United States into Eastern Europe. President Reagan's embargo on allowing Communist countries to buy high technology equipment provides a good market. Other chips are sold to pay narcotics dealers.

In 1979, an employee at INTEL stole 10,000 microcomputer devices worth \$1 million. They passed through black-market distributors in West Germany and eventually arrived at Siemens, a West German computer manufacturer. This robbery was traced when Siemens told INTEL that the untested chips were faulty. That the intensive electronic and human security at Monolithic Memories was bypassed suggests that in this case also the thief was an employee: the company is offering a reward of \$50,000 for information leading to an arrest.

Charlotte K. Beyers

Belgium. The French government intends to build power stations at Chooz, a few kilometres from the Belgian border.

Agreement was reached on a technical proposal to exchange data on sampling and monitoring levels of atmospheric pollution. However, the directive on safety levels for lead in air must wait until a communally acceptable method of monitoring atmospheric pollution is found.

Broad agreement was reached on the European Commission's third environmental action programme and its proposed policy to give environmental considerations greater weight when determining other EEC policies. Finally, despite much discussion, there was no decision to implement the Washington Convention — the EEC regulation on the trade in endangered species is, in fact, reported to be much tougher than the convention.

Jasper Becker

Animal welfare

Promise delayed

The new European legislation to protect laboratory animals seems to be in the hands of legislative snails. Optimists had expected agreement on the draft of a Council of Europe convention on "the protection of animals used for experimental purposes" last May. But the deadline passed, and the hope now is for agreement in March 1982.

Discussions have been under way since 1978, when the Council of Europe appointed an expert committee to draft model legislation. But the experts have so far failed to agree among themselves. Their latest disagreements relate to the "pain condition" that if an animal suffers severe and enduring pain, the experiment must stop and the animal be killed; the use of animals for more than one experiment; and the experimental procedures which should be included in the statistics on animal experiments.

Most of the representatives from the council's 21 member states object to the lack of provision for exemption from the pain condition in the latest draft. The exception is Britain, which wrote the draft when it took over chairmanship of the committee in April 1980. Britain argues that its own legislation, which includes the pain condition without exemption, has worked well since it was enacted in 1929.

Approval of the convention is also being held up by West Germany's objection to a clause which permits an animal used in one painful experiment to be used again. Most other countries accept the clause, some less willingly than others. A further stumbling block relates to whether animals used as controls, for example, should be included in statistics on animal experiments. Sweden believes that they should, but most other countries think their inclusion would push up the apparent number of experiments unnecessarily.

The aim is to resolve these differences before the next annual meeting of the expert committee in March. When agreement is reached, however, a council of ministers will have to approve the draft before it can be laid before individual states for signature and ratification. Because ratification is a lengthy process, the convention is unlikely to come into force for several years.

Meanwhile, the European animal welfare lobby, increasingly vociferous, is asking for tighter controls than those in the latest draft. The Council of Europe's parliamentary assembly, which has no power but which can be influential, has agreed to discuss the lobby's demands and consider recommending changes to the terms of reference of the expert committee. If the assembly decides to recommend changes and the council of ministers agrees to implement them, the convention could be delayed indefinitely or even killed.

That, however, is unlikely. Most states

are eager for a convention as soon as possible, if only to head off the animal lobby's more extreme demands. Some, such as the United Kingdom, have already delayed national legislation pending a convention. Britain's Conservative government now has little more than two years in which to fulfill its last election promise to update the Cruelty to Animals Act of 1876. Two recent private members' bills to do just that have already come to grief. **Judy Redfearn**

French reorganization

Sector structure

Grenoble

Jean-Pierre Chevènement, Minister for Science and Technology in France, has finally decided on the structure and organization of his ministry. He has re-accumulated all the instruments and organizations of science and technology policy — dispersed since de Gaulle last had a major science ministry — and sorted them into a new structure.

The Délégation-Générale à la Recherche Scientifique et Technique, which acted as the relatively small office of the previous science minister, is to be disbanded and its staff re-absorbed. The same fate awaits the Délégation à l'Innovation et à la Technologie (which previously worked within the Ministry of Industry).

The new structure is claimed to be something midway between the highly dispersed "Rothschild" system of the United Kingdom in which individual ministries have control of large sums of research money which they can spend with research councils broadly as they wish, and the highly centralized German system; but the Chevènement structure is on the face of it considerably closer to the German than the British model.

The ministry will be divided into three sectors (plus the cabinet of personal advisers, which will be extended with the addition of a section to evaluate the progress of his policies). There will be a "Direction de la Politique Générale", which, broadly speaking, will administer basic science through the "grands organismes" (the Centre National de la Recherche etc); a "Direction du Développement Scientifique et Technique et de l'Innovation" which will actively promote contacts between science and industry and mastermind the scientific renovation of the French economy (the matter closest to Chevènement's heart); and a high-flying "Mission Scientifique et Technique". This body, headed by Yves Farges, a one-time solid-state physicist and the principal advocate for a European synchrotron radiation source, will act as a kind of interface between the other two sectors and the minister himself.

The staff of the ministry at present numbers some 250. By the time the changes are complete — and the ministry has moved to a new location on the Montagne Ste-

Geneviève (the old site of the Ecole Polytechnique) the staff will probably have grown to 400. **Robert Walgate**

Scientific fraud

In Bristol now

A further case of falsification in the scientific literature has come to light. The following letter has been received from Dr M. J. Purves of the Department of Physiology at the University of Bristol:

SIR — I very much regret to have to report that data published in the proceedings of the 28th International Congress of Physiological Sciences (Purves, M.J. 1981 Cerebral Blood Flow and Metabolism in the Sheep Fetus. *Advances in Physiological Sciences* 9: 199; 126. Pergamon, Oxford & Adademiai Kiado, Budapest) are false. I must also emphasize that none of my colleagues was involved in the preparation of this paper and the responsibility was mine alone.

Bristol, UK

M.J. PURVES

Dr Purves, a reader at the University of Bristol, resigned his post with effect from 1 November after an internal investigation.

The paper concerned describes the use of recently developed techniques for investigating the function of the mammalian (sheep) brain *in utero*. The article describes the use of 5-deoxyglucose as an index of metabolic activity in central nervous metabolism. One of the objectives of the study was to demonstrate that 5-deoxyglucose (not metabolized in the usual way) is taken up more slowly by the fetal mammalian brain during periods when the embryo is asleep.

The University of Bristol seems to have acted quickly since some of Dr Purves's junior colleagues drew attention to the irreproducible features of his published paper earlier in the year. An agreed announcement of the circumstances was delayed for family reasons. Dr Purves said earlier this week that the falsifications consist of the data published. His senior colleagues say that they cannot understand why such a talented person, well-supported by the Wellcome Foundation and the Public Health Services, should have followed such a course.

Erratum

Nature must apologise to one or other (or both) of Sir Andrew Huxley (President of the Royal Society) and Professor T.R.E. Southwood that the latter's photograph appeared last week in place of the former's. The Royal Society's anniversary celebrations this year appear to have been especially prone to accident — the President of the Fellowship of Engineering, half-way through an impassioned speech on the importance of engineering, referred to the present President's predecessor-but-one, Dr A.L. Hodgkin, as Professor Hodgkinson.

Primitive life in France

Paris

The life and times of Europe's earliest inhabitants form the focus of a new exhibition at the Musée de l'Homme in Paris. The exhibition, which opened on 9 December and will run until April 1983, illustrates major advances in human development over more than a million years of European prehistory.

Evidence of human presence beginning with what are reputed to be the oldest stone tools in Europe is traced up to the appearance of fully evolved Neanderthals some 125,000 years ago. Remains from more than 100 sites in 13 European countries are on view, combined with replica living floors and two reconstructions of palaeolithic shelters.

The main advance portrayed in the exhibition, according to Professor Henry de Lumley, exhibition organizer and director of the Laboratory of Prehistory and the museum, is the domestication of fire in Europe some 400,000 years ago. By 300,000 years ago distinct cultures throughout Europe possessed knowledge of fire, were making various tool assemblages, living in organized encampments and hunting in groups.

By this time too, early Europeans had apparently begun to construct more elaborate living shelters with several distinct areas. The remains of two such living floors from sites excavated by Professor de Lumley provide the basis for the striking hypothetical recreations of a hut from Terra Amata and a cave shelter at Le Lazaret (both near Nice in France).



The Terra Amata hut is built from wood poles and pine branches on a replica rocky Mediterranean shore and is based on remains dated at about 380,000 years ago. The Le Lazaret cave shelter suggests an even more comfortable lifestyle by about 130,000 years ago.

Specialists may well debate the interpretation of the stones, bones, dates and reconstructions in this exhibition, but all visitors will go away with a lasting impression of how prehistorians recreate the past and what daily life might have been like for Europe's earliest inhabitants.

Richard Dreiman

CORRESPONDENCE

Psychological slip

SIR — Reviewing H.B. Gibson's biography of Hans Eysenck (*Nature* 5 November, p.44) P.E. Bryant writes: "He (Eysenck), more than anyone, shaped the development of clinical psychology in this country".

This statement may be misleading to a readership as diverse as *Nature's*. In this context the term "clinical psychology" is used, or perhaps linguistically corrupted, to signify the application of behaviouristic or learning theories to the treatment of psychological disorders. It does not refer to the application of psychological theories collectively to the treatment of such disorders, which is the rational inference.

It may well be the case that Eysenck has shaped the development of behaviouristic clinical psychology in this country. However, it would clearly be incorrect to suggest that he has had much influence on the development of certain other principal approaches to clinical psychology such as psychoanalysis.

This duality of meaning is a potential source of considerable confusion, but one which can be simply obviated by qualifying the specific school of clinical psychology referred to.

GEORGE FIELDMAN

Department of Physiology,
Chelsea College, London SW3

Species difference

SIR — There is a serious misunderstanding which is impeding any resolution to the current debate on macroevolution. Recent analysts of the fossil record report a pattern which they claim "strongly supports the notion that speciation is a qualitatively different phenomenon from gradual, intraspecific microevolutionary change"¹. This is true. Others point out that the observed details of fossil history "do not force us to change our views on the genetic mechanisms of the origin of species"². This is also true. The misunderstanding lies in the idea that they are referring to the same thing, for they are not. They are addressing speciation on two entirely different levels.

The model of evolution in contention is the "punctuated equilibrium" proposal of Eldredge and Gould³. Evolution at the specific level is seen as occurring in short rapid bursts followed by relatively long periods of evolutionary non-change or stasis. The words "short" and "long" are used in the context of geological time. That is, morphological changes associated with the transition from an old species to a new one may occur on the order of 10^3 to 10^4 years, which is recorded as an instant in most fossil assemblages. Those who experiment on the evolutionary genetics of living organisms are quite correct in maintaining that this is more than enough time for classical selection mechanisms to have conceivably produced the observed changes. This much is clear to proponents of punctuated equilibrium as well. Why then the call for qualitatively new explanations?

The mechanisms depicting "how" new

species arise have in the past been considered explanations of "why" they develop as well. New species are classically described as being the logical result of long, slow, continuous selective pressure. This continual change is the view of phyletic gradualism. If, however, the prevailing pattern is one of stable equilibrium only punctuated with brief periods of evolutionary change then the mechanism of that change does not tell the whole story. The phenomenon of stasis, not the sudden appearance of new species, is the true source of the conflict with the traditional evolutionary school of thought.

Matters are not reconciled by appealing to selective pressures varying dramatically over time. The environment is dynamic and in a constant state of flux. While some periods may be characterized by a less stable climate than others, there are always marginal environments and isolated communities within which evolution can work. In addition the fossil record does not support the contention that novel stable species arise and replace parental populations only at times of the most extraordinary environmental instability. That is not to say that new species do not arise at times of climatic shifts, but other shifts of equal magnitude may not result in the establishment of new species while novel forms may invade from a marginal locale and become dominant at times of overall relative environmental tranquillity. Thus classical Darwinian theory does not provide an explanation for the punctuated pattern of speciation now emerging from the fossil record.

The models of population genetics may be perfectly adequate at the fine level of how new genetic entities are formed but they do not address questions at the grosser level of the overall patterns. New qualitatively different schemes are therefore needed. Conversely, while theories of a qualitatively different sort are indeed required by the recent reevaluations of the fossil record, the classical genetic mechanisms of speciation are nonetheless compatible with punctual equilibrium.

CLIFF TABIN

Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

1. Williamson, P.G. *Nature* 293, 437–443 (1981).
2. Jones, J.S. *Nature* 293, 427–428 (1981).
3. Eldredge, N. & Gould, S.J. *Models in Paleobiology* (ed. Schopf, T.J.M.) (Freeman, San Francisco, 1972).

Question of faiths

SIR — It is admirable to see scientists such as A.H. Batten and E.W. MacKie (*Nature* 30 July, p.402) come to the aid of the theory of evolution when they perceive that it is threatened by creationism. Such enthusiasm shows their dedication to rational thought. However, when in their eagerness they instead misrepresent its basic scientific tenets one has to wonder just whose side they are on.

Batten is quite correct in criticizing neo-Darwinian evolutionists for being overly dogmatic and thus in fact hastening the rise of the creationist movement. I cannot accept, however, the assertion that the theory of

evolution (that is, descent with modification) is the same as the theory of natural selection and, therefore, if chance mutations are found to be insufficient agents of phylogenetic change one must look for a purposive element to explain the observed hierarchical order in nature. This self-imposed dichotomy leads Batten to reject Popper's critical rationalism for Kuhn's paradigm of "normal science".

We are thus asked to console ourselves with knowing that if we wish to avoid a dogmatic assertion of evolution as a fact then it is necessary to see it as simply the best explanation we have at the moment. Since we must choose between neo-Darwinism and creationism there is not much chance of ever escaping the bonds of the existing paradigm. Batten's complacency with "normal science" and inability to consider other, natural causative agents of evolutionary change (of which the literature is full) only serve to weaken any intended appeal to reason. The arguments presented are also a serious misrepresentation of Popper's philosophy.

MacKie, on the other hand, has chosen to ignore Popper or any other proponent of the hypothetico-deductive method altogether (ignorance of this concept is difficult to believe) and take us almost 400 years into the past. The notion of science as inductive and mythology as deductive might have impressed Francis Bacon, but it has no place in modern thought. Someone should tell MacKie about David Hume. The only way out of the conundrum is to realize that science arises from grand, explanatory ideas, ranging from the metaphysical to the conjectural, and it is only after this initial synthesis that rigid testing is invoked. It is the creationists' uncritical appraisal of their own theories, in both testing procedures and methods of formulation, that marks their endeavours as unscientific. The inductive-deductive barrier has nothing to do with it.

Both authors should learn to live with Popper's observation that rational thought is based on irrational faith in reason. I would add that evolutionary biology is similarly based on a faith that the natural causes we observe are indeed natural causes, and not the manifestations of a supernatural being. We are, then, faced with two faiths: those of rationality and irrationality. The former we can modify with critical argument, the latter we cannot. Yet, just because rationality arises from an irrational faith, we are not compelled to consider every aspiring scientific argument on equal grounds. It is not a case of the "anything goes" attitude of relativism. I believe that is what MacKie worried about as egalitarianism. There should be no problem, since every hypothesis must prove itself in its explanatory power, testability and predictive success. Thus, as MacKie notes, the acceptance procedure for hypotheses must be shamelessly elitist (with respect to the particular hypothesis, not its proponent). Anything less would be uncritical and therefore unscientific. Faith *per se* is no threat to science; dogma and woolly-mindedness are.

RICHARD T. O'GRADY

Department of Zoology,
University of British Columbia,
Vancouver, Canada

NEWS AND VIEWS

Has corticotropin-releasing factor finally been found?

from George Fink

THE precise mechanism of the neural control of the pituitary–adrenocortical system has been the holy grail of neuroendocrinologists for the major part of this century, and it is small wonder that the recent paper in *Science*¹ by Wylie Vale, Joachim Speiss and Catherine and Jean Rivier (and see also ref. 2) on the characterization of a corticotropin-releasing factor (CRF) has stimulated excitement around the various neuroendocrine round tables. The search for the mechanism was begun by Harvey Cushing, Hans Selye and Geoffrey Harris who demonstrated, respectively, that the adrenal cortex is controlled by the pituitary gland, that glucocorticoid secretion is a physiologically important response to stress and that the neural control of the pituitary–adrenocortical system is mediated by chemical substances released from nerve terminals in the median eminence into hypophyseal portal vessel blood. The characterization of CRF was central to a sophisticated understanding of the neural mechanism that subserves the endocrine response to stress, and the clinical significance of the latter, together with the availability of assays for corticotropin (ACTH), led workers in the early 1950s to investigate the nature of CRF before that of any of the other hypothalamic–pituitary regulatory factors. The search was soon confounded by the finding that at least three hypothalamic substances could stimulate ACTH release: catecholamines^{1–4}, vasopressin^{1,5,6} and a substance(s), presumed to be a peptide, distinct from vasopressin.

The systematic search for the nature of CRF in hypothalamic tissue began with the work of Saffran and Schally⁷ and Guillemin and Rosenberg⁸. Schally subsequently joined Guillemin at Houston and, using extracts of neurohypophyseal tissue and mainly gel filtration chromatography, they found CRF activity in three separate fractions which they labelled α_1 , α_2 (similar to α -MSH) and β (similar to vasopressin)^{9,10}. Guillemin *et al.*¹¹ also discovered that extracts of pig hypothalamus contained ACTH and α - and β -MSH, and although remarkable in

that it anticipated by 16 years the discovery of pro-opiomelanocortin in brain¹², this finding increased further the difficulty of interpreting the assay data, and serious work on the isolation of CRF was brought to a halt.

After the isolation of thyrotropin-releasing hormone (TRH), gonadotropin-releasing hormone (GnRH) and somatostatin in the early 1970s, work on CRF isolation began again and has now led to the characterization by Vale *et al.* of a polypeptide CRF from ovine hypothalamus. But is the 41-residue peptide (H-Ser-Gln-Glu-Pro-Pro-Ile-Ser-Leu-Asp-Leu-Thr-Phe-His-Leu-Leu-Arg-Glu-Val-Leu-Glu-Met-Thr-Lys-Ala-Asp-Glu-Leu-Ala-Gln-Gln-Ala-His-Ser-Asn-Arg-Lys-Leu-Leu-Asp-Ile-Ala-NH₂) the physiological CRF? The peptide was isolated in Vale's Peptide Biology laboratory at the Salk Institute from a side fraction of 490,000 hypothalamic fragments that had been used for the isolation of GnRH in the Laboratories for Neuroendocrinology at the Salk Institute, which are headed by Vale's erstwhile mentor and friend, Roger Guillemin. The characterization involved many chromatographic steps, ending in some very neat HPLC, followed by Edman degradation using the spinning cup sequencing modification¹³. The peptide was then synthesized by solid phase. The synthetic and native peptides, at the picomolar to nanomolar range, stimulated the release of ACTH and β -endorphin from primary cultures of anterior pituitary cells, but the synthetic peptide was ten times more potent than the native peptide — the difference was attributed to oxidation of Met 21 of the native peptide to methionine sulfoxide during extraction or purification. The synthetic peptide also stimulated ACTH and β -endorphin in several *in vivo* preparations at dosages (30–3,000 ng per kg) comparable with those at which the other known hypothalamic–pituitary regulatory peptides show activity. Unlike the 40-residue peptide sauvagine that has been

isolated from frog (*Phylomedusa sauvagei*) skin and with which CRF-41 shares several homologous regions, CRF-41 did not stimulate thyrotropin (TSH) or growth hormone release from pituitary cells. The molecular weight of CRF-41 is comparable with that of the CRF activity of porcine hypothalamus which was recently reported by Andrew Schally and associates¹⁴ to elute at about 5,000 daltons. Schally *et al.*¹⁴ speculate that the unexpectedly large-molecular-weight CRF may represent a precursor of a smaller species, but from Vale's data changes to the molecule (assuming that the porcine is similar to the ovine peptide) would have to be confined to a small region at the N terminus. Thus the free acid, CRF(1–39)NH₂ and CRF(10–41)NH₂ were all less than 0.1 per cent as potent as CRF(1–41)NH₂, but CRF(4–41)NH₂ and N-acetyl-CRF were fully active¹.

Vale *et al.* state that it is premature to refer to CRF-41 as *the* CRF, corticoliberin or corticotropin-releasing hormone, and this caution is probably justified because, apart from the previous findings of CRF activities with different chromatographic properties^{1,9,10,14}, there are two points in Vale's paper that cannot be ignored. First, one of the initial steps in extraction of the starting material consisted of partition chromatography. Both the upper and lower phases showed CRF activity, but only the lower phases were available for further purification. Second, when the lower phases were subjected to gel filtration (Sephadex G-50), two distinct zones of the eluate contained CRF activity: zone 1 which eluted close to the void volume (1.3 Ve/Vo) and zone 2 which was retained well beyond the void volume (2.0 Ve/Vo). The 41-residue CRF was isolated from zone 1, while the active fraction of zone 2 had the amino acid composition of [Arg⁸]vasopressin. Thus we are still left with three possible factors that could mediate the neural control of ACTH release — the CRF of the upper partition phases that were not available to Vale, the 41-residue CRF and vasopressin.

However, notwithstanding these remaining uncertainties, the discovery by Vale *et al.*¹¹ is likely to constitute a

George Fink is Director of the MRC Brain Metabolism Unit, Department of Pharmacology, University of Edinburgh.

quantum leap in our understanding of the endocrine response to stress because one more, and possibly the last and most important, factor has been isolated. The deal will be clinched if CRF-41 can be demonstrated in hypophysial portal vessel blood. In fact, early studies on hypophysial portal blood collected from the dog¹⁵ and the rat¹⁶, involving extraction and either Cohn fractionation or ultrafiltration, suggested that CRF activity was associated with a relatively large molecule, perhaps as large as 10,000 molecular weight, in the rat. Apart from measurements of CRF in portal blood, antisera to CRF-41 will be used to demonstrate by immunohistochemistry the location and projections of CRF neurones, and such data, together with those already available on vasopressin-containing neurones, are likely to show why and how the two peptides interact in the hypothalamic-pituitary-adrenocortical response to stress. Numerous studies^{17,18} have shown that vasopressin potentiates hypothalamic CRF activity (and vice versa), and there is no reason to suppose that the control of ACTH release is any less complicated than that of the other two pituitary 'stress hormones', prolactin and growth hormone, which are controlled by at least two, and in the case of prolactin, possibly three (dopamine, a specific prolactin-releasing factor and TRH) hypothalamic transmitters.

Physiology aside, the pharmacological and pharmaceutical importance of CRF-41, like that of any other newly discovered biologically active peptide, is potentially quite enormous. The likely scenario of investigations will be obvious to those who have followed the development of the other hypothalamic peptides. Apart from tests to determine its value as a neuroendocrine tool in man, the effects of CRF-41 on non-endocrine target systems will also be examined. In this respect Vale

et al. have already shown that, like sauvagine and the structurally related urotensin I (extracted from teleost urophypophysis), CRF-41 can significantly reduce blood pressure in the rat. We can also expect a spate of papers on the behavioural effects of CRF-41, and can predict that the peptide will be shown, at

least initially, to affect everything from pole jumping in rats to schizophrenia and Alzheimer's disease. Only after the dust has settled will we know whether Vale and his colleagues have indeed found or at least sighted the holy grail of neuroendocrinology and whether it was worth the long and painful haul. □

Protein-lipid interactions: do the spectroscopists now agree?

from Anthony Watts

THE controversy over whether or not integral proteins in membranes restrict the acyl chain motion of neighbouring lipids appears to have been resolved. Singer and Nicholson's 'fluid mosaic' model implies a uniform lipid environment throughout the membrane, and NMR studies of lipid dynamics seemed to confirm this view. However, ESR experiments reveal two motionally distinct lipid populations in membranes, the motionally slower environment being induced by the large, integral membrane proteins. To explain the ESR observations, the rather controversial but descriptive terms like 'immobilized', 'annular' and 'boundary' lipid were introduced, with subsequent widespread misinterpretation. At a recent meeting*, called to review some of the current ideas about both the spectroscopic and non-spectroscopic aspects of membrane research, a consistent picture of lipid-protein interactions in membranes emerged in which the NMR and ESR observations may be seen to agree by a consideration of the dynamic time scales of the two spectroscopic techniques.

Proton NMR studies of the lipids in reconstituted sarcoplasmic reticulum Ca^{2+} -ATPase membranes were described by Deese *et al.*¹, and in common with phosphorus² and deuterium³ experiments, all the membrane lipids were 'visible' in the NMR spectrum. This result implies that no lipids are grossly restricted in their motion by the Ca^{2+} -ATPase, since immobilized lipids would be undetectable by NMR and the observed spectral intensity would thereby be reduced. All the membrane lipids therefore appear to have very similar motional behaviour, and lipids near the protein interface are almost indistinguishable from the remaining membrane lipids. The ESR technique uses nitroxide lipid spin-labels to probe the lipid acyl chain motion in membranes. It was reported that in the presence of large, integral proteins

like cytochrome oxidase⁴, frog and bovine rhodopsin in the photoreceptor disks^{4,8}, $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ ⁴, Ca^{2+} -ATPase^{6,7} and the acetylcholine receptor^{4,8}, the ESR method resolves two motionally distinct environments for the lipid acyl chains in membranes. One spectral component is characteristic of a label probing a fluid, non-protein-containing lipid bilayer. The other broader component is typical of a spin-label undergoing slow motion and its proportion of the total spectral intensity is often linearly dependent upon the protein concentration in the membrane, which is indicative of a phenomenon not caused by gross protein aggregation and lipid trapping^{4,6}. Thus the idea that a certain fraction of the membrane lipid chains can be immobilized around a protein annulus or boundary has been proposed⁹, an idea which appears at first sight to be inconsistent with the NMR results.

Like any spectroscopic technique, both ESR and NMR have their own dynamic time scales, which for magnetic resonance depends on the rate of motion required to average out the intrinsic magnetization with the applied magnetic field. In the NMR experiment, a nucleus must have a correlation time of faster than around 10^{-4} s to give a spectrum, and if in two distinct environments, an exchange rate of faster than 10^4s^{-1} to give a single-component, averaged spectrum. To explain the ESR observations, the exchange rate of labelled lipid chains between the protein boundary and remaining lipids would have to be slower than $5 \times 10^7\text{s}^{-1}$ to give the impression of two distinct lipid environments¹¹. At the same time, motion in the interfacial region is slow on the ESR time scale, with correlation times of less than 10^{-8} s, but faster than this in the rest of the membrane. Lipid-lipid exchange takes place at a rate of about 10^6s^{-1} (ref.10), and so lipid exchange into or out of the boundary region need only be ten times slower than in the unperturbed, protein-free bilayer (that is, an exchange rate of

- Vale, W., Spiess, J., Rivier, C. & Rivier, J. *Science* **213**, 1394-1397 (1981).
- Spiess, J., Rivier, J., Rivier, C. & Vale, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6517 (1981).
- McDermott, W.V., Fry, E.G., Brobeck, J.R. & Long, C.N.H. *Proc. Soc. exp. Biol. Med.* **73**, 609-610 (1950).
- McDermott, W.V., Fry, E.G., Brobeck, J.R. & Long, C.N.H. *Yale J. biol. Med.* **23**, 52-66 (1950).
- McCann, S.M. & Brobeck, J.R. *Proc. Soc. exp. Biol. Med.* **87**, 318-324 (1954).
- Martini, L., De Poli, A., Pecile, A., Saito, S. & Tani, F. *J. Endocr.* **19**, 164-173 (1959).
- Saffran, M. & Schally, A.V. *Can. J. Biochem. Physiol.* **33**, 408-415 (1955).
- Guillemin, R. & Rosenberg, B. *Endocrinology* **57**, 599-607 (1955).
- Schally, A.V., Andersen, R.N., Lipscomb, H.S., Long, J.M. & Guillemin, R. *Nature* **188**, 1192-1193 (1960).
- Schally, A.V., Lipscomb, H.S. & Guillemin, R. *Endocrinology* **71**, 161-173 (1962).
- Guillemin, R., Schally, A.V., Lipscomb, H.S., Andersen, R.N. & Long, J.M. *Endocrinology* **70**, 471-477 (1962).
- Krieger, D.T. & Liotta, A.S. *Science* **205**, 366-372 (1979).
- Wittman-Liebold, B., Graffunder, H. & Kohls, H. *Analyt. Biochem.* **75**, 621-633 (1976).
- Schally, A.V. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5197-5201 (1981).
- Porter, J.C. & Rumsfeld, H.W. Jr *Endocrinology* **64**, 948-954 (1959).
- Fink, G., Smith, J.R. & Tibballs, J. *Nature* **230**, 467-468 (1971).
- Yates, F.E. *et al. Endocrinology* **88**, 3-15 (1971).
- Gillies, G. & Lowry, P.J. *J. Endocr.* **84**, 65-73 (1980).

* A discussion meeting on 'Protein-lipid Interactions in Membranes' was held in at Airlie House, Virginia on 3-6 October 1981. The text, edited by V. Adrian Parsegian, will appear as Vol. 37, No.1, January 1982 of the *Biophysical Journal*. Page numbers in the reference list refer to the page in the study book.

Anthony Watts is in the Biochemistry Department, University of Oxford.

10^5s^{-1}) to account for both the NMR and ESR observations^{11,12}.

Reconstitution studies allow the experimentalist to manipulate the magnitude of the lipid-protein interaction by varying the relative concentrations of lipid and protein in the membrane. NMR studies of reconstituted membranes were described which indicate that as the protein concentration increases, moderate (about 20 per cent) decreases occur both in the average order (through deuterium order parameter measurements³) and in the rate of lipid chain motion (from T_1 relaxation measurements in proton¹ and deuterium³ studies). An analysis of both the ESR spectral components shows that the rate of lipid chain motion for boundary and non-boundary lipids is decreased by the protein, when compared with the non-protein-containing membrane. Unfortunately, no conclusive information on the amplitude of chain motion can be deduced from these spectra². ESR and NMR observations therefore imply that the 'rough' protein interface both disorders the acyl chains interacting with it and restricts their motion. The magnitude of disordering or motional inhibition again depends on the spectroscopic method being used. These data appear consistent with the idea that proteins induce a rather specific perturbation on the membrane lipids, disordering and motional restriction, unlike the effect of temperature which decreases both the rate and amplitude (together often called 'fluidity') of acyl chain motion.

At the meeting, the relevance of dynamic lipid-protein interactions was questioned during a discussion period by the interjection "So what. . .?". During this and subsequent lively interchanges, it was suggested that lipids at the protein interface (1) may be important for efficient sealing of the bilayer to the protein, particularly vital in transport processes; (2) provide a motional buffer between the peptide backbone to the bulk membrane lipids; (3) sterically maintain at equilibrium the active conformation of a membrane-bound enzyme, which would require a fast motional interaction on the biochemical time scale defined by the catalytic turnover rate of 10^2 – 10^3s^{-1} ; and (4) are able to contort and conform to the protein interface to solvate it in a way that minimizes the energy configuration within the whole membrane. □

New low temperature phases of H_2 and D_2 at ultrahigh pressure

from J.C. Raich and R.D. Etters

At very low temperatures and atmospheric pressure the molecules in solid parahydrogen ($p\text{-H}_2$) and orthodeuterium ($o\text{-D}_2$) are in spherically symmetric rotational states. At a sufficiently high density it is expected that the interactions between the hydrogen molecules deform the average molecular density distributions so that the molecules spontaneously align in an orientationally ordered state. At this transition, the essentially free molecular rotations change to hindered rotations or librations about equilibrium orientations. These equilibrium orientations depend on the crystal structure of the high pressure phases which probably differ from the hexagonal close-packed low pressure structure. This rotation-libration transition in the hydrogens was first predicted by Raich and Etters¹ using a mean field theory. Subsequently mean field calculations² and Monte Carlo simulations³ have supported the first-order nature of the transitions as well as the approximate density dependence. These theories predict critical pressures at roughly 75 kbar for $o\text{-D}_2$ and 270 kbar for $p\text{-H}_2$ using an extrapolated equation of state⁴.

Recent improvements in high pressure techniques, made possible by the development of diamond anvil cells, have renewed interest in the rotation-libration transition and the resulting prediction of new low temperature, ultrahigh pressure, orientationally ordered phases of the solid hydrogens. Indirect evidence of such a transition in hydrogen has recently been reported from the pressure dependence of Raman measurements of the H-H vibrational stretching band $Q_1(1)$ in hydrogen by Sharma, Mao and Bell⁴. The notion that changes in the molecular state of solid hydrogen begin to occur at approximately 200 kbar is supported by the flattening of the curve of the frequency shift $\Delta\nu[Q_1(1)]$ versus pressure which begins to occur at 200 kbar. The decrease in the half-width of $Q_1(1)$ at about 175 kbar is taken as additional evidence for a lessening of molecular rotation. The further flattening of $\Delta\nu[Q_1(1)]$ at higher pressures can be interpreted as an indication that molecular rotation has ceased for pressures exceeding 375 kbar. Sharma *et al.*⁴ also observed that bands attributed to molecular rotation become diffuse and disappear well below 375 kbar at 298 K. However, to attribute the observed changes in the vibrational mode to changes in the rotational motion is only one of several possible explanations. Another possibility might be a pressure-induced change in the electronic charge distribution of the molecule.

Recently, Lagendijk and Silvera⁵ have

studied the rotation-libration transition, taking a different theoretical approach. They investigated the predicted behaviour of the collective rotational excitations in the solid as a function of pressure, and found that increasing the density of the solid leads to a softening of the collective rotational excitations with rotational quantum number $J = 2$. The excitation energy of these $J = 2$ rotons approaches zero at the X boundary point of the Brillouin zone of the face-centred cubic structure for which the calculations were performed. The existence of this roton instability requires a symmetry breaking of the orientational ground state with $J = 0$, resulting in an ordered structure in which the orientational alignment occurs along the $\langle 111 \rangle$ directions of the face-centred cubic lattice. This approach predicts roughly the same transition densities as the earlier mean field treatments. On the other hand, recent Monte Carlo calculations by Aviram, Goshen and Thieberger, using the full anisotropic interaction between H_2 molecules, seem to predict transition densities which are about 60 per cent higher than the mean field densities⁶.

Direct experimental evidence of the existence of a new high pressure phase of $o\text{-D}_2$ and the precursor of a similar phase of $p\text{-H}_2$ has recently been reported by Silvera and Wijngaarden⁷ using Raman scattering which has proved to be a very effective probe of the solid hydrogens. In the low pressure hexagonal close-packed phase, the Raman transitions of interest are three distinct $J = 0 \rightarrow J = 2$ transitions. In the predicted high pressure phase, the crystal structure is probably different, with the result that the corresponding rotational excitation spectrum will be changed significantly. In $o\text{-D}_2$ at about 150 kbar, the $J = 2$ rotons begin to broaden and turn down in frequency. Between 200 and 278 kbar the roton spectrum coalesces into a single broad rotational band. Above 278 kbar the rotational band splits abruptly. These changes are accompanied by a sharp shift of the $\nu = 0 \rightarrow \nu = 1$ or $Q_1(0)$ vibrational frequency. The same change in behaviour is observed when the temperature is lowered below the critical temperature. These changes are considered to be the manifestation of a phase transition to an ordered phase at a temperature $T = 5\text{K}$ and pressures $p \geq 278 \pm 5$ kbar in $o\text{-D}_2$. Although no strong characteristic of a first-order transition was observed, it was not possible to rule out a

J.C. Raich and R.D. Etters are in the College of Natural Sciences, Colorado State University.

1. Deese, A.J., Dratz, E.A., Hymel, L. & Fleischer, S.F. *Biophys. J.* 37, 181 (1982).
2. McLaughlin, A.C. *et al. Biophys. J.* 37, 45 (1982).
3. Fleischer, S., Hymel, L., Tamm, L. & Seelig, J. *Biophys. J.* 37, 47 (1982).
4. Marsh, D., Watts, A., Knowles, P.F., Pates, R.D. & Marsh, D. *Biophys. J.* 37, 215 (1982).
5. Watts, A., Volotovskii, I.D., Pates, R.D. & Marsh, D. *Biophys. J.* 37, 85 (1982).
6. Thomas, D., Bigelow, D.J., Squier, T.C. & Hildago, C. *Biophys. J.* 37, 187 (1982).
7. McIntyre, O. *et al. Biophys. J.* 37, 49 (1982).
8. McNamee, M.G., Ellena, J.F. & Dalziel, A.W. *Biophys. J.* 37, 93 (1982).
9. Jost, P.C., Griffith, O.H., Capaldi, R.A. & Vanderkooi, G. *Proc. natn. Acad. Sci. U.S.A.* 70, 480 (1973).
10. Trauble, H. & Sackmann, E. *Chem. Phys. Lipids* 7, 324 (1971).
11. Knowles, P.F., Watts, A. & Marsh, D. *Biochemistry* 18, 4480 (1979).
12. Jost, P.C. & Griffith, O.H. *Ann. N.Y. Acad. Sci.* 348, 391 (1980).

simultaneous hexagonal close-packed to face-centred cubic structural change analogous to the case of the zero-pressure o -H₂ and p -D₂ orientational transition.

In p -H₂ for $p > 500$ kbar, the solid appears to be approaching the region where a similar phase transition might occur. However, higher densities seem still to be required. It is likely, however, that the rotation-libration transition in p -H₂ will occur before the metal-insulator transition.

The observed transition pressures, 278 kbar for o -D₂ and in excess of 500 kbar for

p -H₂, are much larger than the theoretical predictions. Whether this difference is due to the approximations made in the various theoretical approaches, or to an inaccurate representation of the intermolecular potential, remains to be elucidated. In particular, the use of asymptotic forms for the anisotropic contributions to the H₂-H₂ interaction potential at high pressures should be questioned seriously. A comparison of the results of Silvera and Wijngaarden⁸ with those of Sharma, Mao and Bell may be difficult as the latter started with fluid hydrogen at room

temperature and the former with p -H₂ at only a few degrees Kelvin. □

1. Raich, J.C. & Etters, R.D. *J. Low Temp. Phys.* **6**, 229 (1972).
2. England, W., Raich, J.C. & Etters, R.C. *J. Low Temp. Phys.* **22**, 213 (1976); Felsteiner, J. & Friedman, Z. *Phys. Rev. B* **8**, 3996 (1973).
3. Aviram, I., Goshen, S. & Thieberger, R. *J. Low Temp. Phys.* **38**, 371 (1980).
4. Silvera, I.F. & Goldman, V.V. *J. chem. Phys.* **69**, 4209 (1978).
5. Sharma, S.K., Mao, H.K. & Bell, P.M. *Phys. Rev. Lett.* **44**, L886 (1980).
6. Lagendijk, A. & Silvera, I.F. (to be published).
7. Aviram, I., Goshen, S. & Thieberger, R. *8th AIRAPT Conf.*, Uppsala, Sweden (1981).
8. Silvera, I.F. & Wijngaarden, R.F. *Phys. Rev. Lett.* **47**, L39 (1981).

New pathways for chlorinated dioxins

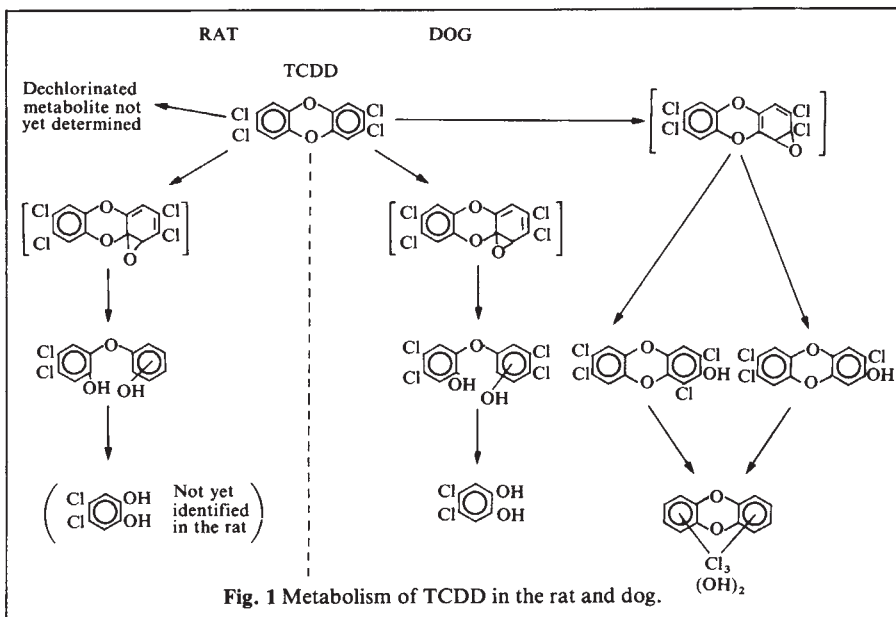
from Alistair Hay

UNTIL recently many would have argued that 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) — a toxic contaminant produced in the manufacture of 2,4,5-trichlorophenol — was not metabolized in animals. The chemical seemed to require no metabolic activation to make it more toxic and did not appear to be metabolized before excretion.

But times have changed, and so has the evidence. It is still true that TCDD is active in its own right, but it is equally clear that some metabolism of the chemical occurs in animals. Poiger and Schlatter (*Nature* **281**, 706; 1979) were the first to report that TCDD was hydroxylated in the rat. Others are reporting similar findings. At a recent workshop* R.A. Neal (US Chemical Industry Institute of Toxicology) reported that Golden Syrian hamster hepatocytes will metabolize TCDD *in vitro*, producing a pattern of metabolites similar to those excreted in the bile and urine of the animal.

The faeces (bile) is the major route of excretion in the rat, hamster and mouse. The TCDD which is retained in tissues after administration appears to be unmetabolized. For the small amount of the chemical which is metabolized — generally of the order of one per cent — the process appears to begin with hydroxylation to phenols. The hydroxyl groups, Neal suggests, are then converted to glucuronides prior to excretion.

Before this reaction scheme can be proved, the metabolites will have to be positively identified. H. Poiger and H.R. Buser (University of Zurich and Swiss Federal Research Station respectively) have identified metabolites in the bile of the rat and the dog which enables them to suggest that in these two animals, TCDD is metabolized by the reaction scheme shown in Fig.1. They claim that metabolism of TCDD is a process of detoxification. Extracting the metabolites from the bile of dogs and testing for toxicity in the guinea



pig shows them to be at least 100 times less toxic than the TCDD itself.

According to S. Safe (Texas A & M University) there are common features in the metabolism of polychlorinated dibenzodioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and polychlorinated biphenyls (PCBs). With all these chemicals the rate of metabolism decreases as the degree of chlorination increases; and the major metabolites formed are phenolic compounds which are generally less toxic than their parent hydrocarbon precursors.

When aromatic hydrocarbons reach a certain size and lipophilicity it seems that they are poorly retained in the body. There is considerable species variation in the metabolism of different isomers of these compounds and lipid stores have a major role in the regulation. H.B. Matthews (NIEHS, Research Triangle Park, US) reported that animals eventually reach a state of dynamic equilibrium with adipose and other tissues playing their part in sustaining blood TCDD concentrations. Should the blood concentration fall, dioxin will move out of other tissues and back into

the blood in sufficient amounts to restore the equilibrium. The process, says Matthews, is not rapid, but occurs over a period of days, perhaps even as long as a week.

When animals are starved TCDD is mobilized from adipose tissue and this could lead to blood dioxin levels reaching concentrations which ultimately prove toxic. Fortunately, in human subjects this situation is unlikely to occur. The body burden for TCDD in humans is thought to be low. But low as blood levels of this chemical are, it is becoming clear that it can stay around for a long time.

C. Rappe (University of Umea, Sweden) reported that pentachlorodibenzofurans could be detected in the liver of a young Japanese man — at a level of 1.2 ng g⁻¹ — 12 years after he was exposed to the chemicals. The PCDFs were present as contaminants in PCBs which affected a batch of cooking oil in Japan in 1968. The

*The 'International Symposium on Chlorinated Dioxins and Related Compounds' was held at the Sheraton National Hotel, Arlington, Virginia, 25-29 October 1981.

Alistair Hay is a lecturer in the Department of Chemical Pathology, University of Leeds.

victims affected by this oil are known as the 'Yusho I' cases. A similar 'Yusho II' event occurred in Taiwan in 1979.

It is currently standard practice to assess exposure to chlorophenols by measuring the levels of these chemicals in urine. Rappe, however, suggests that blood measurement of the PCDF and PCDD contaminants present in the chlorinated phenols is a superior method. Urine concentrations of chlorinated phenols fall rapidly to undetectable levels in a very short time whereas PCDDs and PCDFs can be detected in blood years after exposure. Although the significance of TCDD levels in blood is not yet clear, many believe that it is important to have more of these measurements to try and relate blood levels to clinical symptoms.

It is a pity this information is not available now for assessing the carcinogenic risk from TCDD. Reports from Sweden (Hardell and Sandstrom *Br. J. Cancer* 39, 711; 1979) suggest that forestry workers exposed to chlorinated phenols and the herbicide 2,4,5-T have a sixfold higher incidence of soft tissue sarcomas. A repeat of the investigation by one of the Swedish scientists to take account of some criticisms of the earlier study confirms the initial findings (Hardell *Scand. J. Work envir. Hlth* 7, 119; 1981).

However, studies conducted in other countries have so far failed to confirm Hardell's reports. A cohort of herbicide applicators in Finland exposed to phenoxy herbicides for at least two weeks between 1955 and 1971 have been monitored since 1972. The incidence of tumours in this group is no higher than would be expected and no cases of soft tissue sarcoma have been observed. Similarly, preliminary findings from a study in New Zealand, reported by A.H. Smith (Wellington Hospital), in which soft tissue sarcoma cases were compared with patients with other forms of cancer showed that the individual's place of work bore no relationship to the type of tumour seen.

The different findings from the various studies are disturbing. So much so that a workshop study group on the health implications of TCDD exposure concluded that it was a matter of urgency that additional studies be done to determine whether there is indeed a causal relationship between soft tissue sarcoma and exposure to phenoxy herbicides such as 2,4,5-T.

Experimental design of the epidemiological studies may have some bearing on the different results obtained. If so, it is important that new studies should take account of this. There is general agreement that all the studies performed to assess the carcinogenic risk from exposure to TCDD and phenoxy herbicides were well designed. However, it is felt that some studies may have bias which has so far gone unrecognized. If this is the case it will be some time yet before there is unequivocal evidence about the health risk from exposure to 2,4,5-T. □

Parasites affect behaviour of mice

from F.E.G. Cox

PARASITIC WORMS are ubiquitous in small mammals, even in many laboratory colonies, and are almost invariably ignored in behavioural studies. Two recent investigations indicate, however, that they may markedly influence both the acquisition of dominance and general exploratory behaviour of mice.

When male mice are reared in isolation and then put together fighting occurs and a hierarchy of dominance is established with one dominant male. This simple procedure has allowed an Australian scientist, W.J. Freeland, to investigate the effects of nematode infections on the behaviour of mice¹. Using groups of two infected and one uninfected mice, he found that those given 50 or 150 larvae of *Nematospiroides dubius* became dominant as often as uninfected ones but those given 250 larvae became dominant less frequently than would be expected. Once dominance in uninfected mice was established, however, a subsequent infection with 250 larvae did not lead to its loss. Freeland argues that, as dominance may be related to mating success, infected mice are at a reproductive disadvantage and that females mating with uninfected or lightly infected males protect themselves and their offspring from infection and also select males with genes for resistance to infection.

Even though these results are convincing and the conclusions self-evident, they are subject to a number of reservations. First, the number of mice used in each trial, three, was small, although 20–30 trials were used for each level of larval infestation; a more satisfactory procedure would have been to use 6–10 mice. Second, the numbers of worms that brought about behavioural changes were relatively large. In natural infections the majority of mice carry small numbers of worms and very few carry large numbers² and 250 is more than would normally be expected. The reasoning also fails to take into account the true nature of nematode infections in mice in the wild for the female is as likely to become infected from another male as from her mate as there is no permanent pair-bonding and, as rodent territories are not absolute, males tend to roam over large areas.

The suggestion that female mice might effectively select for genes for resistance is also open to question. Although the importance of genetic factors in resistance to nematodes in general and to *N. dubius* in particular is well documented³, individual animals acquire their initial infections with large and small numbers of worms purely by the

chance ingestion of infective stages and only later mount an immune response; so in primary infections, like the ones described, the female would be selecting for luck rather than resistance. Furthermore, more active males range further and thus have more opportunities to become infected; this is well known for ectoparasites such as ticks⁴ and is presumably also true for worms of various kinds.

The second set of experiments were carried out in the United States at the State University of New York and Cornell⁵. In these experiments, mice were infected with the dog nematode *Toxocara canis*. The larvae of this worm cause considerable damage by burrowing about the body of abnormal hosts and may cause blindness in children. In mice, the larvae may enter the nervous system and brain. The American workers infected mice with *T. canis* and subjected them to a number of behavioural tests to assess such activities as exploratory behaviour, areas traversed, avoidance reactions and spatial discrimination in mazes. The various tests were carried out sequentially over a period of 115 days and in all tests infected mice did less well than uninfected ones. The authors conclude that *T. canis* infection reduces exploratory activity, learning and motor co-ordination. These results are easier to understand than those of the dominance experiments for the mice infected with *T. canis* were all found to have parasites in the brain at autopsy, but a direct correlation between brain damage and behavioural changes has not yet been established.

These two sets of experiments taken together, however, do suggest that parasites can modify the behaviour of rodents and although this is well known in other hosts the gross behavioural changes are usually brought about by physical damage such as is caused when a parasite lodges in the eye or brain or occupies a considerable amount of the body cavity. The interactions between parasite and host may, however, be more subtle than this and parasites in the gut or kidney may affect the odours produced by the host or may even affect its hormone balance. It is likely that parasites will be increasingly implicated in alterations in host behaviour; however the nature of these changes will not be resolved by the use of simple behavioural experiments but will have to follow the advances being made in our understanding of the underlying mechanisms of animal behaviour. □

1. Freeland, W.J. *Science* 213, 461 (1981).

2. Crofton, H.D. *Parasitology* 62, 179 (1971).

3. Wakelin, D. *Adv. Parasit.* 16, 219 (1978).

4. Randolph, S.E. *J. Anim. Ecol.* 44, 451 (1975).

5. Dolinsky, Z.S. *et al.* 213, 1142 (1981).

F.E.G. Cox is Professor of Zoology at King's College, London.

Useful tropical legumes

from Robert M. May

AMONG the many potential applications of research on recombinant DNA that are exciting commercial enthusiasm is the hope that cereal and other crops may be genetically engineered to fix their own nitrogen. This is a consummation devoutly to be wished, because current trends in the cost of energy are making nitrogenous fertilizers extremely expensive for peasant farmers. At the same time, however, there appear to be many plants in the family Leguminosae that are greatly under-exploited; some are extensively used in one locality but unheard of elsewhere, while others are virtually unknown but have particular attributes that suggest they could be developed into major crops. The legumes provide their own nitrogenous fertilizer through bacteria that live in nodules on the roots of the plant, where the bacteria convert nitrogen gas from the air into soluble compounds the plant can absorb and use. Although a few other plant families include species with this ability, legumes produce the bulk of biologically fixed nitrogen. Even today, the total nitrogen added to the world's soil by cultivated leguminous crops exceeds that added by fertilizers.

Leguminous plants are found throughout the world (including a few aquatic legumes), but the greatest variety is in the tropics and subtropics. There are approximately 650 genera and 18,000 species of Leguminosae, which makes them the third largest family of flowering plants (after the Compositae and the Orchidaceae). The family is divided into three subfamilies: the Caesalpinioideae, comprising about 2,800 species which are mainly trees of the tropical savannas and forests of Africa, South America and Asia; the Mimosoideae, again comprising about 2,800 species, which are mostly small trees and shrubs of semi-arid tropical and subtropical regions of Africa, North and South America, and Australia (*Acacia* species are prominent in this category); and the Papilionoideae, comprising around 12,000 species, which are mainly herbs and are distributed worldwide. Legumes can most easily be recognized by the pods they bear; these pods come in an enormous variety of sizes and shapes, from one not much bigger than a pinhead to one the size of a tennis ball, and they contain from one to several dozen seeds.

A recent study (*Tropical Legumes: Resources for the Future*, National Academy of Sciences, Washington, DC, 1981), emphasizes that "Of the thousands of known legume species, less than 20 are used extensively today". Cultivation is most commonly for the seeds (called beans, grain legumes, or 'pulses'), which rank

second only to cereals as a source of human and other animal food. Nutritionally, leguminous seeds are typically two to three times richer in protein than are cereals. Before the advent of the potato in Europe, beans constituted much of the diet of the poorer classes. Today, they remain a major food in Latin America, on the Indian sub-continent (especially lentils and chickpeas) and in the Far East (especially soybeans). Greater emphasis on pulses, in contrast to the 'Green Revolution's' emphasis on cereals, could do much to alleviate the chronic protein deficiency found in many developing countries.

The NAS report notes many promising pulses that stand in need of further research. The virtues of the Winged Bean have been extolled earlier (*Nature* 266; 590, 1977). The Bambara Groundnut is an African pulse which seems to constitute a complete food: although "its seeds have less oil and protein than peanuts, they do have more carbohydrates and make a well balanced food with a calorific value equal to that of a high-quality cereal grain. . . it can thrive in arid inferior soils where peanuts fail, it resists pests and diseases, and, if managed well, can give high yields". Another example is the Moth Bean, which is "an exceptionally hardy South Asian legume that thrives in hot, dry, tropical conditions, [producing] nutritious seeds and green pods, leafy forage for hay or pasture, and a soil-building 'living mulch' to complement orchard crops and to protect and improve fallow land. . . It is likely to prove very useful in extending agriculture production into marginal regions — especially those bordering tropical arid zones". An interesting candidate for a cash crop is the Yeeb, from the Horn of Africa, which flourishes in tropical arid zones and produces seeds with taste akin to macadamia and pistachio nuts.

In earlier centuries, when natural dyes were the only way to colour fabrics, legumes played a key part in the history of Western exploration and colonial development. The blue dye indigo, extracted from small leguminous shrubs of the genus *Indigofera*, was a major product of India, traded eastwards to China and westwards to Europe. More valued than spices, it was a major reason for establishing Portuguese, Dutch and British colonies in India. In South America, the small leguminous tree *Caesalpinia echinata* yielded a wine-red essentially identical to

brazil (from *Caesalpinia sappan*), a dye that Europe had been importing from Asia for centuries. This dye soon became the most important export from the new land to Portugal, whence the merchants became known as *bresilheiros*, and the region as *Brasil* or *Brazil*.

Turning from the past to the future, the NAS report systematically lists many kinds of uses to which legumes may be put, beyond the relatively familiar pulses.

Native tribes in several rural areas of the world eat tubers harvested from wild or cultivated legumes. Most of the species so used have not been subjected to scientific research, but a few that have been analysed show a remarkably high protein content (sometimes several times higher than the root crops in conventional use in the tropics). The NAS report urges the study of these little known root crops, with a view to improving and disseminating the most productive and nutritious ones.

There are several interesting fruits that could be more widely exploited. For instance, Mediterranean peoples have for millennia enjoyed the sugar-rich, mealy pulp of carob pods. "The handsome, drought-tolerant carob tree deserves more research and widespread exploitation in semi-arid areas, for in addition to pulp it provides a chocolate substitute, high-protein flour and an industrial gum, as well as shade, beautification, erosion control and forage." Although in common use for centuries, the tamarind remains largely unimproved and little cultivated; up to half the mass of a tamarind pod is sour-sweet pulp, which is used in refreshing drinks (and in Worcestershire sauce). The honeylocust can be easily grown from seed, suckers, or cuttings, and grows fairly fast to produce giant pods from the fourth or fifth year. The sweet, succulent pulp of the pods is relished by humans and other animals. As I have one in the backyard, I learn with mixed feelings that it is one of the hardiest, most adaptable and most generally useful fodder trees known, but that care is needed in new locations as it can form dense thickets and become a pest (as it has in Queensland).

Too often, woody plants are overlooked in programmes of research for improved forage in tropical regions. Yet, as the NAS panel notes, browse shrubs and trees can complement herbaceous pasture species and can be crucial to the survival of animals during droughts (when shallow-rooted species often shrivel to straw). Emphasizing that many species undoubtedly await discovery and exploitation, the NAS study focuses on the thorny, flat-topped *Acacia tortilis* bush (one of the most drought-tolerant trees of semi-arid areas of Africa and the Middle East, and a great soil stabilizer) and on other species of *Acacia* and *Prosopis* (both of which are drought-resistant and well adapted to light, sandy soils).

Among many fast growing trees of potential commercial significance is the

Robert M. May is Class of 1877 Professor of Zoology at Princeton University.

Leucaena leucocephala (itself the subject of a separate book, *Leucaena leucocephala: New Forage and Tree Crop for the Tropics*, NAS, 1977). Resembling a mimosa, the tree can grow 65 feet tall and 16 inches thick in five years (roughly six times the rate of growth of an oak), giving wood suitable for paper pulp, furniture or building. Bushy varieties of the species help prevent soil erosion, while providing a nitrogenous fertilizer or a cattle feed rich in protein. *Acacia auriculiformis*, a robust and little known tree from New Guinea, can grow with exceptional vigour on problem soils such as eroding hillslopes, mining spoil, laterite and sand, as well as in highly acid and alkaline sites. More generally, the NAS report looks at other fast growing trees (especially of *Acacia* and *Albizia* species) and concludes that "many woody

legumes are pioneer species that colonize newly cleared sites. To outdo the competition, they grow fast and are very precocious and vigorous trees that nodulate well. With the desperate need for reforestation, erosion control, firewood, and paper, as well as other wood products in developing countries, many such leguminous trees are worth widespread testing".

The rosewoods — classic furniture woods renowned for their gorgeous colours, beautiful grain and exceptional technical properties — come from species of *Dalbergia*, slowly growing leguminous trees found throughout the tropical regions of the world. Most species are being harvested to commercial extinction, and silvicultural research has been so neglected that in most cases it just is not known whether

the trees can be cultivated. The fascinating properties of several other slowly growing, 'luxury timber' trees are discussed in the NAS report. Here the problems of lack of information are compounded by the commercial reality that the 'discounted present value' of such slowly growing trees is relatively low, making their cultivation economically unattractive (see *Nature* 263; 91, 1976).

Last, but not least, legumes have given us many of the world's most exquisite flowering plants: wisteria, laburnum, poinciana, orchid trees, cock's comb coral tree, the raintree and many others. The NAS book lists and, better, illustrates the other beautiful leguminous species that should be far more widely known and more extensively planted, before they become extinct. □

Small minded — the characterization of interplanetary dust by electron microscopy

from Colin Pillinger

IN recent years, the extraterrestrial sample community has been made increasingly aware that a third category of material, besides meteorites and lunar rocks, is available for laboratory study. Samples of interplanetary dust may be collected in the Earth's stratosphere by appropriately equipped high flying aircraft. The grains, known as 'Brownlee particles' after Don Brownlee (of the University of Washington, and more lately the California Institute of Technology), who championed the cause of the collection programme, are now being collected in a programme coordinated by NASA from the Johnson Space Center.

Collecting cosmic dust is not a new pastime. Almost a hundred years ago the Challenger expedition recovered spherules of presumed cosmic origin from the slowly accumulating Pacific sediments¹. Brownlee himself obtained copious quantities of extraterrestrial material from this source with his so-called 'cosmic muck rake', a magnetic device trailed behind a ship and capable of processing thousands of tonnes of sediment per day. Similar attempts have been made to find extraterrestrial debris in Arctic and Antarctic snow and a variety of other sediments. Perhaps the first report of micrometeorite debris can be attributed to an occurrence of 'Cryoconite' (Icedust) in Greenland ice² reported in 1872.

Particle collection at high altitudes is possible because the braking effect of the terrestrial atmosphere is able to concentrate the incoming dust flux to around 10^{-3} particles per m^3 . The frictional heating caused by deceleration depends on the size of the particle and its emissivity. Very small grains ($<10 \mu m$) do not reach anything like the melting point of the constituent material as they have been

slowed to terminal velocity from a speed of $25 km s^{-1}$ at an altitude of 100 km (ref. 3). Probably half the grains of size $<10 \mu m$ and with density ~ 1 will not have been heated to greater than $550^\circ C$ for more than two seconds⁴.

That comparatively mild thermal history has meant that very flimsy low-density grains whose structures are often described as 'porous', 'reentrant' or even 'fairy castle' have survived for collection. That they are truly extraterrestrial is shown by rare gas contents and isotope abundance patterns which can only have been produced by exposure to the solar wind^{5,6}. Solar flare particle tracks which might also provide evidence of space exposure have not, however, been found, suggesting that either some annealing took place during entry or the lifetime of interplanetary dust grains at 1 AU is much shorter than anticipated⁶.

Availability of a special sample is one thing, participation in its study is another. Among the few who can sensibly contribute to the analysis of Brownlee particles is the group at the McDonnell Center for Space Sciences of Washington University, St Louis. Since 1976, this group, particularly P. Fraundorf, has been examining 11 particles in the size range $4-15 \mu m$ by analytical techniques involving electron microscopy. Perhaps this five year effort and future studies on the same 11 grains are put in perspective by Fraundorf's analogy that a $5 \mu m$ grain to an electron microscopist is the same as a three ton hand specimen to a field geologist.

Fraundorf's efforts are now directed towards the chemical and mineralogical

taxonomy of interplanetary dust particles and in a recent paper⁷ he attempts to emphasize their differences, not only from known meteorite classes, but from each other. He has now become so familiar with his grains that each has acquired a name, some of which describe the characters of the specimens. Thus, 'Blowup' 'exploded' on crushing into hundreds of silicate fragments of wide-ranging composition each coated with some low atomic number material. 'Lost and Found' was predominantly a crystalline coarse-grained aggregate which was difficult to handle because of electrostatic charging.

Electron microprobe and single grain neutron activation studies have already shown⁸ that Brownlee particles have compositions which may be called 'chondritic', that is, similar to chondritic meteorites and relatively unfractionated compared with solar abundances. Fraundorf has looked more deeply into the compositional variations. One hundred and fifty analyses from regions as small as $10^{-15} cm^3$ from his 11 grains give average elemental abundance ratios which would be described as chondritic but on inter- and intra-grain scale the variations are extensive. In the case of meteorites, element abundance ratios can fairly uniquely define the class to which a sample belongs — none of the eleven grains could be assigned to a distinct chondritic group with any certainty.

Plotted on the standard Urey-Craig plot (reduced iron + sulphide/Si versus oxidized iron/Si), interplanetary dust particles (IDPs) fall in the region ascribed to the most reduced H, L and E chondrites, samples which often show some degree of metamorphism. None of Fraundorf's eleven grains plots with the oxidized carbonaceous chondrites,

Colin Pillinger is in the Department of Earth Sciences, University of Cambridge.

samples generally thought to be the most primitive (least metamorphosed) meteorites, even though numerous arguments exist to show that IDPs are relatively unmetamorphosed and probably contain per cent quantities of carbonaceous phases (low atomic number material).

It would be easy to see how incoming IDPs were oxidized during entry but not reduced; the overall oxidation state of IDPs must reflect their formation in a highly reduced region of the primitive solar nebula. Fraundorf points out from his mineralogical studies that some of the fine-grained magnetite decoration on crystal surfaces might be due to atmospheric heating, but could just as easily fit characteristics predicted by core/mantle models for interstellar grains.

Before Fraundorf's studies, only X-ray powder diffraction had been used to recognize minerals in whole IDPs (ref.9). Now as a result of both single- and polycrystal electron diffraction studies, coupled with dark and bright field imaging and consideration of concurrent chemical analysis data, magnetite, pyroxene, olivine, pyrohotite, taenite and cohenite (the latter as a possible deposition product from a reducing gas phase) have all been identified. Perhaps just as important is that amorphous zones of a particular chemistry (such as S/Fe > 2) have been observed. One surprise is the absence of hydrated silicates, again distinguishing IDPs from the matrix material in carbonaceous chondrites. The possibility that water was lost by exposure in space or atmospheric entry cannot be discounted but the evidence seems more and more that Brownlee particles are distinct from known meteorite classes and really are a third category of extra-terrestrial sample.

It is hoped that soon many other investigators will become small minded and accept the challenge of Brownlee particles. The full characterization of primitive materials and their taxonomic classification is a necessity to our understanding of the early Solar System. It is still largely supposition that IDPs are derived from cometary sources but the indirect evidence to this origin continues to grow. Just the conjectured availability of cometary dust alone should mean a tremendous cross-fertilization between the laboratory-bound analysts and the spacecraft experimenters who hope to receive data from the ESA Giotto mission to Halley in 1986. □

Clues for the origin of cosmic rays

from Peter Meyer

DURING the 1960s and early 1970s it became possible to measure the elemental composition of the cosmic radiation in fine detail. This achievement constituted a major breakthrough towards recognizing the elusive origin of the high energy particles that pervade the Galaxy. In particular, it was found that elements over the entire range of the periodic system are present in the cosmic radiation, that the relative abundance ratios of the cosmic ray elements are strikingly similar to those found in the elements that make up the Solar System. Hoyle first made the suggestion, now generally accepted, that the chemical elements heavier than lithium are produced by nucleosynthesis in the interiors of stars. It seems clear that the cosmic ray particles cannot be primordial, but rather must have been part of stellar material before their acceleration.

The crucial questions remaining to be answered concern the identification of the sites of origin and of the acceleration mechanisms responsible for the high energy particles. From the point of view of energy alone, it is clear that a powerful particle accelerator must be available. The cosmic ray population must be replenished, on the average, every ten million years, which is the currently accepted value of the mean storage time of the particles in the Galaxy before escape. This population is known to have existed with an intensity similar to that now observed throughout the lifetime of the Solar System, that is, about 4,500 million years. Supernova explosions are the most energetic known phenomenon in the galaxy, and were therefore long suspected to be responsible for the origin of cosmic ray and possibly to imprint characteristic

abundance distributions on the ejected nuclei.

Major theoretical work has led to a deeper understanding of various nucleosynthesis processes and in attempts to describe the elemental and isotopic composition of the Solar System, the end products from hydrostatic and explosive nucleosynthesis were investigated. Explosive syntheses are expected to occur on a short time scale in conditions similar to those observed in the supernova phenomenon. Hence, the deviations of cosmic ray elemental and isotopic compositions from the Solar System or interstellar matter abundance distributions provide information that reveals the origin of the particles.

Two recent publications in *Astrophysical Journal Letters* (1 August 1981) address this question in two different ways. One paper deals with the isotopic abundance of several elements of low atomic number, the other with the elemental abundances of nuclei with atomic number between 26 and 40 (Fe to Zr)². The abundance distribution of the elements lying between helium and nickel in atomic number is not a particularly sensitive discriminator between the nucleosynthesis processes forming the cosmic ray elements, but the isotopic composition is. The discovery of an 'anomaly', that is, a deviation from Solar System abundance, in the isotopic composition of the element neon was therefore received with great excitement³⁻⁸.

The most accurate results were obtained from two high-resolution experiments carried out on the spacecraft ISEE-3, for the Caltech group⁵ and for the Berkeley group^{1,8}. They present additional, new values for the abundance ratios ²²Ne/²⁰Ne, ²¹Ne/²⁰Ne, ²⁵Mg/²⁴Mg, ²⁶Mg/²⁴Mg, ²⁹Si/²⁸Si and ³⁰Si/²⁸Si. It is claimed that all these isotopic ratios show enhancements of the neutron-rich isotopes over Solar

Peter Meyer is Professor of Physics at the Enrico Fermi Institute and Department of Physics, University of Chicago.



100 YEARS AGO

THE conduct of competitive examinations in China seems to be farther from perfection than might be expected in the case of such an ancient institution. The *Peking Gazette* contains a memorial from one of the censors complaining that the matsheds which are erected at the entrance to the examination hall in the capital to issue tickets of admission to competitors are frequently overturned by the rush of applicants, that an unseemly crowding

and snatching of tickets from the officials take place, and that candidates break the rule prohibiting them from leaving the compartments in which they are isolated during the examination. They are allowed, he says, to fetch their food themselves (examinations in China last from thirty-six hours to three days at a stretch) from the kitchens, and they meet and converse freely. Prepared essays, the memorialist fears, are passed in from outside during these hours by the student's friends. Again, when the lists of successful candidates are posted up, a tumultuous crowd assembles outside the gates; bands of the unsuccessful ones obstruct the progress of the chief examiner, employing threats and entreaties to prevail on him to alter the lists.

From *Nature* 25, 15 December, 160 (1881).

1. Murray, J. *Nature* 29, 585 (1884).
2. Nordenskiöld, N.A.E. *Geol. Mag.* 9, 355 (1872).
3. Brownlee, D.E., Tomandl, D.A. & Olszewski, E. *Proc. 8th Lunar Sci. Conf.* 149 (1977).
4. Fraundorf, P. *Geophys. Res. Lett.* 10, 765 (1980).
5. Rajan, R.S., et al. *Nature* 267, 133 (1977); Hudson, B., Flynn, G.L., Fraundorf, P., Hohenberg, C.M. & Schirk, J. *Science* 211, 383 (1980).
6. Fraundorf, P., Flynn, G.J., Shirek, J. & Walker, R.M. *Proc. 11th Lunar planet. Sci. Conf.* 1235 (1980).
7. Fraundorf, P. *Geochim. cosmochim. Acta* 45, 915 (1981).
8. Brownlee, D.E. in *Protostars and Planets* (ed. Gehrels, T.) 134 (University of Arizona Press, 1978); Ganpathy, R. & Brownlee, D.E. *Science* 206, 1075 (1979).
9. Brownlee, D.E., Tomandl, D.A. & Olszewski, E. *Proc. 8th Lunar Sci. Conf.* 149-160 (1977).

System abundances. This enhancement amounts to a factor of about 1.6 except for the ratio of 4.1 measured for $^{22}\text{Ne}/^{20}\text{Ne}$. It is not yet understood why the Ne isotopes show a larger anomaly; a theoretical model by Arnett⁹, extended to the special case of these cosmic ray isotopes by Woosley and Weaver¹⁰, would predict all the ratios to be of the same order.

There is no doubt that the isotopic ratios for the above cosmic ray nuclides differ from the standard Solar System values. The critical reviewer of this work remains concerned that the standard Solar System values are always quoted without errors, which are obviously difficult to estimate. A most interesting comparison is made by Wiedenbeck and Greiner with recent spectroscopic determinations of the silicon isotopes in interstellar clouds¹¹. In some interstellar clouds an enhancement of the neutron-rich isotopes ^{29}Si and ^{30}Si is observed which is not dissimilar to that seen in the galactic cosmic rays. If such observations turn out to be representative for the interstellar medium, the anomalous isotopic composition of the cosmic rays remains compatible with their source being pre-existing interstellar material which itself carries those anomalies.

The second sensitive indicator for the nature of the cosmic ray sources is the chemical abundance distribution of the very rare ultraheavy elements with atomic number beyond that of the iron/nickel group. Attempts to measure that abundance distribution date back to the late 1960s, only a few years after the discovery of ultraheavy elements in the cosmic radiation. The early measurements pointed towards an enhancement of elements produced by rapid nucleosynthesis. The work by both Binns *et al.*², carried out on the spacecraft HEAO-3, and Fowler *et al.*¹⁶, on ARIEL-6, has much improved charge resolution. Binns *et al.* conclude that, within the charge range of Z from 26 to 40, the HEAO-3 results are inconsistent with a cosmic ray source that is dominated by r -process nucleosynthesis, but quite

consistent with an elemental abundance distribution similar to that of Solar System material. A crucial element ratio supporting this conclusion is the Se/Sr abundance ratio, found to be about ten times that expected from the r -process.

Clearly, the recent work on isotopic anomalies as well as on the abundance distribution of superheavy elements in the cosmic rays is consistent with the interpretation that the source of the particles may be the interstellar medium. It is important to note that these papers represent interim results, that are expected to be improved upon, and that in both directions of research much remains to be done. Several crucial isotopic ratios ($^{34}\text{S}/^{32}\text{S}$, $^{54}\text{Fe}/^{56}\text{Fe}$, $^{58}\text{Fe}/^{56}\text{Fe}$) must be

determined accurately with instruments having higher sensitivity than exists at present; this work should go hand in hand with new attempts to determine the corresponding isotopic abundances in the interstellar medium. The potentially very revealing elemental distribution of elements with $Z > 40$ must be measured accurately. The Pt/Pb ratio and abundance of the actinides are particularly crucial in identifying the origin of the particles. Some time in the future measurements of isotopic composition may include the ultraheavies; this would provide a powerful tool not only for determining the sources of cosmic rays, but also for understanding the contemporary interstellar medium. □

Middle atmosphere dynamics

from John Gregory

MUCH of our understanding of the dynamics of the middle atmosphere¹⁻⁹ comes from data gathered in the mid-sixties by rocket techniques and by meteor radars. The techniques each had their particular limitations and left the mesosphere and lower thermosphere (60–100 km above the Earth's surface) as the least known regions of the atmosphere. In the 1972 Cospar International Reference Atmosphere², for example, data were insufficient to assemble a latitude-season cross-section of meridional winds, 60–120 km above the Earth's surface, while the zonal cross-section was incomplete for the higher latitudes and altitudes. This situation provided the impetus to set up a co-ordinated program of international scientific research, the Middle Atmosphere Program (MAP).

In the last decade, two new techniques have provided additional data. First, satellite-borne radiometers^{10,11} have improved global coverage of the stratosphere and lower mesosphere, to 80 km, notably with respect to the large-scale temperature structure and its perturbation by planetary waves and stratospheric warmings. Weighting factors broaden with altitude and result in decreasing altitude discrimination and this, coupled with low temporal resolution, means that data for the study of motions such as tides and gravity waves may not be readily available from these techniques. The derivation of geostrophic winds from temperatures, after inversion of radiances, requires knowledge of pressure fields at a lower boundary, such as is available from balloon and rocket sondes.

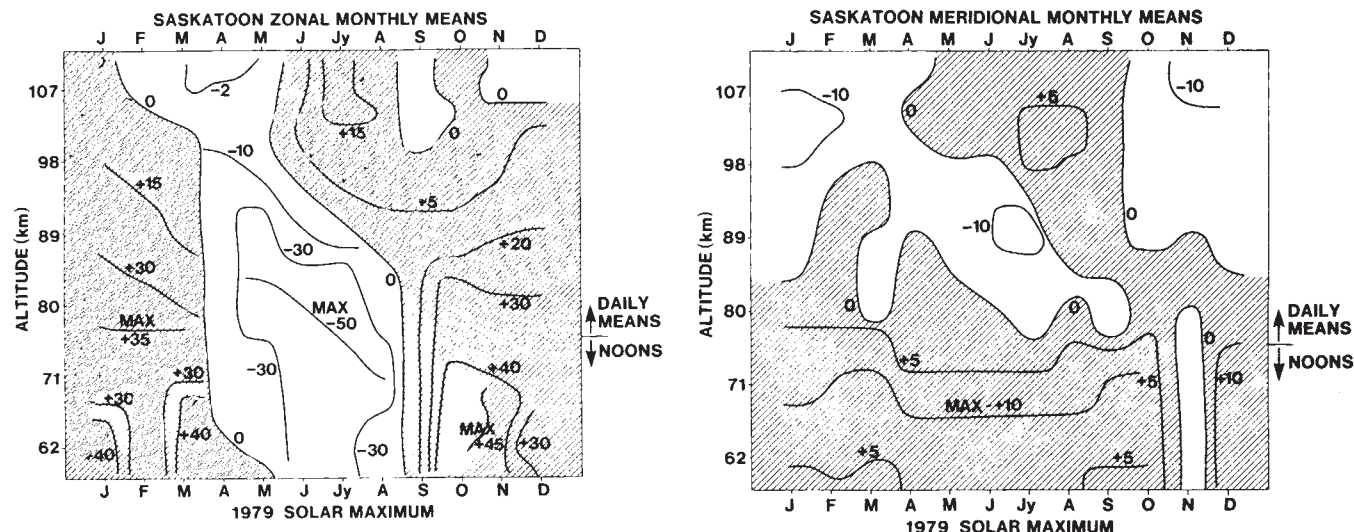
The second technique comprises ground-based radar, in several variants, to which may be allied lidar. The oldest form of radar measurement of winds¹², in which

a moving diffraction pattern at ground level is sampled at spaced antennae, has been applied to MF partial radiowave reflections¹³⁻¹⁵ from irregularities in electron density. Examples of the wind field determined by fully automated version of this method are shown in Figs 1 and 2. A later form¹⁶⁻²⁰, in which Doppler measurements of winds are made at VHF, utilizes reflections from electrons in the mesosphere, and aerosols in the troposphere and lower stratosphere. In the region 30–60 km, irregularities in neutral density provide some scattering, but extremely sensitive radars are required to detect it. These two variants of radar, once conceived as differing in principle, have been shown²¹ to be based on the same scattering irregularities. Thomson scattering from electrons at UHF can be processed to yield winds, 60–100 km²². Lidars^{23,24} are now being employed to derive winds and density profiles, 30–80 km. The altitude resolution of the radar techniques is from 0.2 to 3 km, and is thus adequate for studies of gravity waves and tides. VHF and UHF radars are advantageous for turbulence studies.

Available techniques are already providing data for MAP projects in dynamics, in particular, a coordinated study of the behaviour of the middle atmosphere in winter (MP-1; Chairman, K. Labitzke) and equatorial wave dynamics (MP-2; Chairman, I. Hirota). Study groups have identified observational requirements for dynamical problems — for example, the role of tides, gravity waves and turbulence in respect to momentum and energy budgets (Study Group 3;

John Gregory is Professor of Physics in the Institute of Space and Atmospheric Studies, University of Saskatchewan.

1. Wiedenbeck, M.E. & Greiner, D.E. *Astrophys. J. Lett.* **247**, L119 (1981).
2. Binns, W.R. *et al. Astrophys. J. Lett.* **247**, L115 (1981).
3. Fisher, A.J., Hagen, F.A., Machl, R.C., Ormes, J.F. & Arens, J.F., *Ap. J.* **205**, 938 (1976).
4. Garcia-Munoz, M., Simpson, J.A. & Wefel, J.P. *Astrophys. J. Lett.* **232**, L95 (1979).
5. Mewald, R.A., Spalding, J.D., Stone, E.C. & Vogt, R.E. *Astrophys. J. Lett.* **235**, L95 (1980).
6. Freier, P.S., Young, J.S. & Waddington, C.J. *Astrophys. J. Lett.* **240**, L53 (1980).
7. Dwyer, R. & Myer, P. *Proc. 16th int. Conf. on Cosmic Rays*, Kyoto, Japan, Vol.12, p.97 (1979).
8. Wiedenbeck, M.E. & Greiner, D.E. *Phys. Rev. Lett.* **46**, 682 (1981).
9. Arnett, W.D. *Astrophys. J.* **166**, 153 (1971).
10. Woosley, S.E. & Weaver, T.A. *Astrophys. J.* **243**, 651 (1981).
11. Wolff, R.S. *Astrophys. J.* **242**, 1005 (1980).
12. Fleischer, R.L., Price, P.B., Walker, R.M., Maurette, M. & Morgan, G., *J. geophys. Res.* **72**, 355 (1967).
13. Fowler, P.H., Adams, R.A., Cowan, V.G. & Kidd, J.M. *Proc. R. Soc. A301*, 39 (1967).
14. Shirk, E.K. & Price, P.B. *Astrophys. J.* **220**, 719 (1978).
15. Israel, M.H., Klarmann, J., Love, P.I. & Tueller, J. *Proc. 16th int. Conf. on Cosmic Rays*, Kyoto, Japan, Vol.12, p.69 (1979).
16. Fowler, P.H., Walker, R.N.F., Mashed, M.R.W., Moses, R.T. & Worley, A. *Preprint* (University of Bristol, 1980).



The wind field in the mesosphere and lower thermosphere, as determined by spaced-antenna radar, operating at 2.22 MHz, in a fully-automated mode, at Saskatoon, 52N, 107W. The zonal flow, Fig.1, is in good agreement with European results at similar latitudes, but the meridional flow, Fig.2, shows a poleward component below 80 km in summer, contrary to model calculations.

Chairman, M.A. Geller), also the role of gravity waves and turbulence in respect to transport of minor constituents (Study Group 2; chairman J.D. Mahlman).

Techniques are potentially available for the majority of probable programmes, with two major exceptions. Gravity waves and turbulence are at present amenable to study only by ground-based and rocket techniques so the determination of their global characteristics seems a goal that is unlikely to be met. Also, measurements of mean vertical speeds, which range from cm per second in the troposphere to m per second in the lower thermosphere, are not yet practicable. The two related techniques, Doppler and spaced-antenna radar, are in principle capable of doing so, but there are economic as well as technical limitations. Finally, it may be noted that satellite measurements of winds by Doppler optical and microwave techniques are under development, but mainly due to spacecraft scheduling, are unlikely to produce data for about a decade.

Models of global circulation have been developed from basic equations of motion and knowledge of radiation fluxes and constituent concentrations³⁻⁹. A second approach has been to calculate the meridional circulation, including vertical motion, from observed radiance fields²⁵. So far, only zonal mean motion has been modelled. Most models are able to reproduce the major features of zonal mean flow in the stratosphere and mesosphere, for example, the winter eastward and summer westward flow (see Fig. 1). A major difficulty which seems to have been encountered in nearly all models is the development of excessive Coriolis torques from meridional flow, the latter being derived from the requirement that vertical motion compensate for observed departure from radiative equilibrium, for example, the cold mesopause in summer. In consequence, computed zonal speeds

are generally much in excess of those observed, and modellers have resorted to imposing arbitrary damping. Since planetary waves have now been excluded from the list of possible sinks of momentum in the summer mesosphere²⁶, opinion currently favours tides and gravity waves as the damping mechanisms. It appears that before significant advances in modelling mean flow can be made, observational and theoretical studies of tides and gravity waves, together with their interaction with mean flow, must progress.

Much of the theoretical study and modelling of specific dynamical features, for example, the quasi-biennial oscillation²⁷ and stratospheric warmings^{28,29}, as well as the global circulation, preceded MAP, and would undoubtedly have proceeded independently of it. The purpose of MAP is to facilitate such work, and this is possible in two ways. First, desirable observational programs are identified. The space and time scales appropriate to each dynamical process are determined. Thus, for example, attention has been drawn to the relative lack of observations in the tropical upper mesosphere. Member nations of The Inter-Union Special Committee on Solar Terres-

trial Physics (SCOSTEP) are encouraged to determine what contribution they can make to global observational programmes, or what work they may specifically attempt in their territory. Thus European nations have made special studies of winter middle atmosphere conditions, and are planning more.

Second, the systematic compilation of data is to be attempted. This is intended to permit the assembly of a dynamic climatology of the middle atmosphere, initially between 30 and 60 km, and as data become available, between 60 and 120 km. In addition, the data are intended to form a source for incorporation into models, and a reference for comparison. In particular, such data will, it is hoped, serve to guide aeronomers concerning the transport of trace constituents, in respect to which an 'eddy diffusion coefficient' is currently utilized. It is well appreciated that many of the aeronomic problems of the middle atmosphere, for example, the concentration of water vapour in the summer mesosphere, or of electrons in the winter mesosphere cannot be determined by photochemical studies alone, and a major aim of MAP is to ally these latter to appropriate dynamical studies. □

- Murgatroyd, R.J. *Q. Jl. R. met. Soc.* **83**, 417 (1957).
- Groves, G.V. in *Cospar International Reference Atmosphere* (1972).
- Murgatroyd, R.J. & Singleton, F. *Q. Jl. R. met. Soc.* **87**, 125 (1961).
- Leovy, C.B. *J. Atmos. Sci.* **21**, 327 (1964).
- Cunnold, D.M., Alyea, F., Phillips, N. & Prinn, R. *J. Atmos. Sci.* **32**, 170 (1975).
- Manabe, S., Mahlman, J.H. *J. Atmos. Sci.* **33**, 2185 (1976).
- Schoeberl, M.R. & Strobel, D.F. *J. Atmos. Sci.* **35**, 577 (1978).
- Holton, J.R. & Wehrbein, W.M. *Pageoph.* **118**, 284 (1980).
- Crane, A.J., Haigh, J.D., Pyle, J.A. & Rodgers, C.F. *Pageoph.* **118**, 307 (1980).
- Curtis, P.P., Houghton, J.T., Peskett G.D. & Rodgers, C.D. *Proc. R. Soc. A337*, 135 (1974).
- Gille, J.C. *J. Atmos. terr. Phys.* **41**, 702 (1979).
- Briggs, B.H., Phillips, G.J. & Shinn, D.H. *Proc. phys. Soc.* **B63**, 106 (1950).
- Fraser, G.J. *J. Atmos. terr. Phys.* **30**, 707 (1968).
- Gregory, J.B., Meek, C.E., Manson, A.H. & Stephenson, D.G. *J. appl. Met.* **18**, 682 (1979).
- Vincent, R.A. & Stubbs, T.J. *Planet. Space Sci.* **25**, 441 (1977).
- Woodman, R.F. & Guillen, A. *J. Atmos. Sci.* **31**, 493 (1974).
- Balsley, B.B. & Gage, K.S. *Pageoph.* **118**, 452 (1980).
- Fukao, S., Kato, S., Aso, T., Sasada, M. & Makahira, T. *Radio Sci.* **15**, 225 (1980).
- Röttger, J., Klostermeyer, J., Czechowsky, P., Rüster, R. & Schmidt, G. *Naturwissenschaften* **65**, 285 (1978).
- Royvik, O., Gibbs, K.P. & Bowhill, S.A. *Handbook for MAP 2*, 37 (1981).
- Briggs, B.H. *J. Atmos. terr. Phys.* **42**, 823 (1980).
- Mathews, J.D. *J. geophys. Res.* **B1**, 4671 (1976).
- Clemsha, B.R., Kirchhoff, V.W.J.H., Simonich, D.M. & Batista, P.P. *J. geophys. Res.* **B6**, 868 (1981).
- Houghton, J.T. *Q. Jl Roy. met. Soc.* **103**, 487 (1978).
- Holton, J.R. & Lindzen, R.S. *J. Atmos. Sci.* **29**, 1076 (1972).
- Schoeberl, M.R. *Rev. Geophys. Space Phys.* **16**, 521 (1978).
- Holton, J.R. *A. Rev. Earth planet. Sci.* **8**, 169 (1980).

REVIEW ARTICLE

The early Universe

Edward W. Kolb^{*†} & Michael S. Turner^{†‡}^{*}Theoretical Division, University of California, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA[†]Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA[‡]Astronomy and Astrophysics Center, University of Chicago, Chicago, Illinois 60637, USA

In the past few years one of the most exciting areas of research in physics has been the interdisciplinary field of cosmology and particle physics. The NSF's Institute for Theoretical Physics in Santa Barbara devoted a 6-month program and an intensive 1-week workshop to the subject. A brief review is given of both the workshop and this field which is attracting attention, in part, because the early Universe seems to be the only laboratory in which to study grand unification.

AT the Institute for Theoretical Physics (ITP) on the campus of the University of California at Santa Barbara, the NSF is trying a new approach for tackling problems in theoretical physics. The idea is to bring together for 6-months to 1-year periods people from all over the world who are working on similar problems, in the hope that their collective effort will be more productive than the sum of what the individuals could accomplish working separately. Usually several different workshops run concurrently, and so there is the possibility that connections between apparently disconnected areas of physics will be found.

From January to June 1981, a program was held on the early Universe, which focused on the close relationship between cosmology and particle physics. Simultaneously, there were related programs in the areas of quantum gravity and gauge field theories. Within the past 5 years there has been much exciting work and interesting developments in the interdisciplinary field of particle physics and cosmology. The primary tool of cosmologists has been the large telescope, but now they are interested in deep underground mine experiments built to detect proton decay, in experiments on mountain tops searching for rare cosmic ray events, in experiments performed at accelerators and reactors which are designed to detect axions and neutrino oscillations, in balloon flights looking for antiprotons, in high altitude U-2 flights which carry instruments to measure anisotropies in the 3 K microwave background, in low temperature searches for fractional charge states, in deep underwater muon and neutrino detectors, and even in a proposed monopole search at an iron-ore processing plant.

Traditionally, the large accelerator has been the primary tool of the particle physicist, however, they are now interested in reactor experiments, in deep underground mine experiments, in bulk matter searches for exotic particles left over from the big bang, in the cosmic abundance of helium and deuterium, in the distribution of galaxies, in the structure of the microwave background, and even in experiments designed to detect solar neutrinos. It is not uncommon to find as many cosmologists as particle physicists at a particle physics seminar and vice versa. The connection between the two disciplines is, of course, the big bang. In the early Universe temperatures as high as 10^{32} K were reached, corresponding to average particle energies as high as 10^{19} GeV. The particles present and their interactions determined the early evolution of the Universe, and the type of relics that remain. Particle physicists are constantly seeking higher energies where the world appears simpler, and the early Universe seems to be the only laboratory with large enough energies ($\geq 10^{14}$ GeV) to study models of the unification of the strong, weak and electromagnetic forces. Cosmologists are asking particle physicists if neutrinos have small rest masses, and if the proton is unstable. If neutrinos have small rest masses they may determine the large-scale dynamics of the Universe; and if the proton is unstable, baryon number violating reactions may

explain why the Universe is predominantly matter, rather than having equal amounts of matter and antimatter. Conversely, particle physicists are asking cosmologists about the cosmic environmental impact of exotic particles which either cannot be produced at accelerators or would have escaped detection.

The culmination of the early Universe program at Santa Barbara was a week-long workshop which brought together experimental and theoretical high energy physicists, theoretical astrophysicists, astronomers and cosmologists. The workshop focused on many of the recent exciting developments and the important outstanding problems of common interest in both particle physics and cosmology. We shall now highlight some of the work presented at the meeting, and review the present status of the field of cosmology and particle physics.

Standard big-bang model

It is now generally accepted, and supported by observations that the Universe began from a hot big bang (for review see refs 1-3). The evidence includes: the universal expansion or 'Hubble flow'; the 3K cosmic microwave background; the large abundance of D and ^4He ; and on the theoretical side, the singularity theorems of Hawking and Penrose. In the homogeneous and isotropic Friedmann-Robertson-Walker (FRW) model, the scale factor of the Universe, $R(t)$, is governed by

$$(\dot{R}/R)^2 \equiv H^2 = 8\pi G\rho/3 + \Lambda/3 - k/R^2 \quad (1)$$

where ρ is the total energy density, Λ is a possible cosmological term, and k is the curvature signature, which when $R(t)$ is appropriately scaled is 0 (open and flat), 1 (closed and positively curved), or -1 (open and negatively curved). The energy density of radiation $\propto T^4 \propto R^{-4}$ as $T \propto R^{-1}$, and the energy density of matter $\propto R^{-3}$. For $T \geq 10^5$ (Ωh^2)K ($t \leq 3 \times 10^{10}$ s $\Omega^{-2} h^{-4}$), radiation dominates the energy density and the right-hand side of equation (1), as it increases faster than the other two terms as $R \rightarrow 0$ (Ω is the present ratio of the energy density to the critical energy density, and the Hubble parameter $H = 100h$ km s $^{-1}$ Mpc $^{-1}$). During the radiation epoch, the age and temperature of the Universe are related by

$$t = \left(\frac{45}{16\pi^3 G} \right)^{1/2} g_*^{-1/2} T^{-2} = 2.41 \mu\text{s } g_*^{-1/2} T_{\text{GeV}}^{-2} \quad (2)$$

where g_* is the total number of degrees of freedom of all species which are ultrarelativistic (that is species with $m \ll T$), and throughout units are used such that $\hbar = c = k_B = 1$. For $t \geq t_{\text{Planck}} \sim 10^{-43}$ s, quantum corrections to general relativity should be small, and equations (1) and (2) should be valid. When T is much greater than the mass of a particle, that particle and its antiparticle are about as abundant as the photons. So for $T \geq 10$ GeV ($t \leq 10^{-8}$ s), all the particles which we have produced at accelerators ($e^\pm$, $\mu^\pm$, $\tau^\pm$, $\nu\bar{\nu}$, and all the quarks) were as abundant

as photons. Earlier than $t \sim 10^{-12}$ s, average particle energies were greater than 1,000 GeV—the highest energies achieved at particle accelerators. At $t = t_{\text{Planck}}$, $T \sim 10^{19}$ GeV. It is clear why the early Universe has been called the ultimate particle accelerator.

Elementary particles and interactions

In the past decade or so much progress has been made in the understanding of fundamental particles and their interactions: the strong, weak, and electromagnetic forces (for review see ref. 4). There is now good evidence that the fundamental particles of nature are quarks—five different types are known to exist, and a sixth type is also believed to exist (u, d, c, s, b and t), and leptons—six different types are known (e^- , ν_e , μ^- , ν_μ , τ^- and ν_τ). Quarks and leptons apparently come in pairs: (u, d) and (ν_e , e^-); (c, s) and (ν_μ , μ^-); (t, b) and (ν_τ , τ^-). Each quark also comes in three colours. It is now believed that each of the fundamental forces can be described by a gauge theory, in which the interactions are mediated by gauge bosons. Each gauge theory is distinguished by an invariance group, usually unitary (symbol $U(n)$) or special unitary (symbol $SU(n)$), whose elements are 'rotations' on an n -dimensional complex vector. Each gauge theory has its particular group (and value of n) which distinguishes it from all others: it has particular algebraic relations among its charges, and among the gauge fields that emanate from them, which are determined by the group—which thus determines the physical nature of the fields described.

For example, the gauge theory of electromagnetism is a $U(1)$ theory, and the gauge field is the electromagnetic field whose particle representation is the photon. The colour force between quarks is described by an $SU(3)$ gauge theory. The gauge fields are the colour fields, and their particle representations are called gluons. The strong forces between nucleons and other baryons and mesons are believed to be a residual effect of the colour force, analogous to van der Waals' forces. The weak and electromagnetic interactions can be described by one unified theory, the $SU(2) \times U(1)$ gauge theory. (Here each element of the direct product gauge group is composed by taking the product of one element from $U(1)$ and another from $SU(2)$.) In this theory the gauge particles are the photon, $W^\pm$ and Z bosons. Normally these bosons, like all primitive gauge bosons, would be massless and long range, but in the Weinberg–Salam model which uses the $SU(2) \times U(1)$ gauge group an additional concept is introduced: that the vacuum is not empty but uniformly filled with certain scalar fields, the Higgs fields, which possess a subset of the charges of the gauge group and slow the propagation of—or give mass to—the corresponding subset of gauge bosons. Here these are taken to be $W^\pm$ and Z^0 , which are given masses of around 80 GeV. This process called 'spontaneous symmetry breaking' has a characteristic energy scale below which the gauge symmetry is broken (and the relevant gauge bosons get mass) and above which the full symmetry is restored (and all the gauge bosons are massless). (The usual analogy used for spontaneous symmetry breaking is ferromagnetism. Maxwell's equations are rotationally invariant; however, below the Curie temperature the rotational invariance of a ferromagnet is spontaneously broken when the magnetization chooses a specific direction.)

Encouraged by the success of the $SU(2) \times U(1)$ unified gauge theory, particle theorists have speculated that all three forces are unified under one gauge group (not yet determined) which is spontaneously broken. In the simplest class of models, the breaking scale has been calculated to be $\sim 10^{14}$ GeV (ref. 5). The unification of the three interactions requires new interactions, interactions which violate baryon and lepton number conservation. These two conservation laws were never very pleasing; unlike charge or energy and momentum conservation, these quantities do not couple to long-range forces.

The startling prediction of grand unified theories (GUTs) at the instability of the proton, with a lifetime of $0(10^{30})$ yr. At least five large underground experiments designed to detect proton decay are under construction, and several smaller experiments

are operating (for review see ref. 6). A Japanese–Indian collaboration in the Kolar Gold Fields has recently reported three candidate events for proton decay, translating into a proton lifetime of $\sim 3 \times 10^{30}$ yr (ref. 7).

The simplest picture of unification predicts no new interactions between 300 GeV and 10^{14} GeV; however, there very well may be new interactions yet to be discovered. There is, of course, the problem of also unifying gravity—'superunification'. Grand unification seems to occur ~ 4 orders of magnitude lower than the scale at which gravity should be as strong as the other forces ($\sim 10^{19}$ GeV), and it is hoped that gravity can be unified independently, at a later stage. At present, various efforts are being made to obtain a quantum theory of gravity, and even unify gravity (for example, supergravity theories).

In the gauge theory description the world becomes simpler at higher energies—full symmetry is restored, and the forces are predicted to become weaker, so that the Universe, during its earliest stages when particle number densities were as high as 10^{99} cm^{-3} , can be treated as an ideal gas of weakly interacting point-like particles. Clearly the early Universe is the only laboratory in which the ultrahigh energies that particle physicists are now interested in were achieved. Unfortunately, this laboratory shut down over 1,000 Myr ago, so that one must search for fossils which remain from the earliest epochs. The fossil record to date includes: the existence of structure in the Universe (galaxies, clusters, and so on), the baryon-to-photon ratio $\cong (3-5) \times 10^{-10}$, the abundance of D and ^4He , and the 3K microwave background and its angular structure, the absence of a large cosmological term or curvature term (that is, both are less than, or of the order of the matter density term), the possible existence of fractionally charged states, and the absence of various exotic species that might have been present if various proposed particle physics theories were correct.

Primordial nucleosynthesis

One of the great triumphs of the hot big-bang model is primordial nucleosynthesis (for review see ref. 8). During the period $\sim 1-200$ s after the 'bang' when T was $\sim 1-0.1$ MeV, $\sim 25\%$ of the mass of the Universe was synthesized into ^4He , with lesser amounts of D, ^3He and ^7Li being produced. The nucleosynthetic yields depend on nuclear reaction rates which are measured in the lab, η (the ratio of nucleons to photons), and the expansion rate of the Universe which itself is determined by g_* , the number of effectively relativistic species present [see equation (2)]. By using the observed abundance of ^4He , which is the sum of the primordial and stellar produced ^4He , big-bang nucleosynthesis has been used as a very powerful probe of this early epoch. Specifically, one can set an upper limit to the number of light (< 1 MeV) neutrino species, as they would have been present and contributed to g_* during this epoch. However, to do this, one needs a lower bound to η . If one assumes that baryons (and not, for example, neutrinos) dominate the mass of small groups of galaxies, then dynamical studies of these systems provide such a bound, and the limit of ≤ 4 neutrino species has been obtained. In quark–lepton symmetric theories this implies ≤ 8 quark flavours⁹.

At the Santa Barbara workshop Steigman reported new work¹⁰ which uses the D and ^3He abundance to obtain a lower bound to η , and also results in the limit of ≤ 4 neutrino species, independent of whether or not baryons dominate the mass of small groups of galaxies. Note that this lower bound is consistent with that obtained from studies of small groups of galaxies, and so is evidence that small groups are dominated by baryons. Studies of rich clusters of galaxies result in a much higher lower bound to η , indicating that they are not predominantly baryons, but perhaps neutrinos. In addition, when ^4He is used to obtain an upper limit to η , η is restricted to the range $3-5 \times 10^{-10}$, or $\Omega_b \cong (0.01-0.02) h^{-2} (T_c/2.7 \text{ K})^3$. Thus, nucleons alone cannot close the Universe. With this new lower bound, the standard model makes the firm prediction that the primordial ^4He production is 0.23–0.25. The primordial mass fraction of ^4He must be less than the presently observed fraction of ^4He . ^4He

observations range from 0.23 to 0.32 with some of the values having stated errors as small as ± 0.01 or 0.02. The higher values can be accounted for by stellar production (and indeed usually involve systems with normal metal abundances, indicating that stellar processing has occurred). If the primordial mass fraction of ${}^4\text{He}$ is found to be less than 0.23, then the standard model is in trouble¹¹.

If this should occur there are several alternatives: a cold big bang in which the 3 K background must be produced astrophysically^{12,13}; perhaps some of the primordial D and ${}^3\text{He}$ (or ${}^4\text{He}$) was destroyed by astrophysical processes; or a large lepton number might be hidden in the undetectable neutrino seas. In the standard model of primordial nucleosynthesis it is assumed that the lepton number-to-photon ratio, like η , is $\ll 1$. Reeves discussed primordial nucleosynthesis when the lepton number-to-photon ratio is of $O(1)$ (ref. 14). Relaxing the assumption that the lepton number is small allows much more freedom, and it is possible to explain simultaneously a low primordial ${}^4\text{He}$ abundance, and the observed abundances of D and ${}^3\text{He}$. Although baryosynthesis usually results in lepton and baryon numbers of the same order of magnitude, one of us (E.W.K.) discussed a GUTs scenario which results in $\eta \sim 10^{-10}$ and a large lepton number¹⁵.

Baryosynthesis

Although the laws of physics are almost exactly matter-antimatter symmetrical at the microscopic level (the only exception is a small effect in the neutral kaon system), the Universe appears to be grossly asymmetrical. There is no evidence for the existence of appreciable amounts of antimatter in our Galaxy. If antimatter exists in the Universe in non-negligible amounts, then the size of matter or antimatter domains must be greater than the size of clusters of galaxies¹⁶. The matter asymmetry is usually quantified as the baryon number, or more exactly the baryon-to-photon ratio $B = (n_b - n_{\bar{b}})/n_\gamma \approx \eta$, where n_b , $n_{\bar{b}}$, and n_γ are, respectively, the number densities of baryons, antibaryons and photons. (For review of baryosynthesis, see refs 17, 18.) The Universe today has $B = 3-5 \times 10^{-10}$; for $t \leq 10^{-6}$ s after the 'bang' ($T \geq 1$ GeV), baryons and antibaryons were as abundant as photons, and this ratio then translated into a matter-antimatter fractional asymmetry of ~ 1 part in 10^{10} . Perhaps the most exciting application of GUTs to cosmology is the possible explanation of the baryon asymmetry. Sakharov¹⁹ discussed the conditions necessary for an initially symmetrical universe, $n_b = n_{\bar{b}}$, to evolve into an asymmetrical universe, $n_b \neq n_{\bar{b}}$. The three ingredients outlined by Sakharov are: baryon number nonconserving interactions; C- and CP-violating interactions, that is, interactions which violate matter-antimatter symmetry; and a departure from thermal equilibrium. The necessity of baryon number violation and interactions that distinguish between particles and antiparticles (CP-violation) is clear. As the entropy of a system is a maximum when the baryon number vanishes, a departure from thermal equilibrium is also necessary for the microscopic interactions to produce the asymmetry.

In GUTs quarks and leptons are usually grouped together in a unified manner, and so there is a gauge equivalence between leptons and quarks. Physically, this gauge equivalence corresponds to interactions mediated by gauge bosons (which, of course, violate baryon number conservation). As C- and CP-violations are observed in the kaon system, it is not unreasonable to expect such violations to occur also in baryon number violating reactions. The expansion of the Universe can provide the final condition: if the expansion is rapid enough, then the temperature of the universe ($T \sim R^{-1}$) changes too fast for particle interactions to maintain equilibrium. All three of these ingredients are incorporated into the out-of-equilibrium, drift and decay (British) model proposed by Weinberg and Wilczek^{20,21}.

The generation of the baryon asymmetry has motivated many people to take the idea of unification seriously enough to study various grand unified models in detail. In fact, until proton decay

is seen, the existence of a cosmological baryon asymmetry is the only 'experimental' evidence for baryon number violation. Several detailed calculations of the generation of the baryon number were presented at the conference.

Although the standard scenario for the development of the baryon asymmetry has rapidly gained wide acceptance, there are still basic uncertainties in the model. One uncertainty involves the assumption of thermal equilibrium as an initial condition. The effect of non-equilibrium initial conditions was discussed by one of us (M.S.T.)²² who showed that the magnitude of the asymmetry produced can be different if the Universe was not initially in a state of thermal equilibrium. Another uncertainty is the magnitude and type of CP-violation. It seems almost certain that the baryon asymmetry cannot be directly related to the CP-violation in the K^0 - $\bar{K}^0$ system. Nanopoulos²³ presented a model that related the CP-violation responsible for the baryon asymmetry to the neutron electric dipole moment, and not to the kaon system. (The electric dipole moment of the neutron is another possible manifestation of CP-violation.) The calculations suggest that the neutron electric dipole moment should be just smaller than the present upper bound.

Although many details need to be worked out, there is a feeling that a solution to the fundamental cosmological problem of the origin of the baryon asymmetry is at hand. A possible problem is the detection of low-energy cosmic ray antiprotons by Buffington *et al.*²⁴ discussed at the conference by Gaisser. Although the antiprotons ($\bar{p}$) detected are small in number (≈ 14) and small in proportion to the protons detected ($\bar{p}/p \approx 10^{-4}$), and although $\bar{p}$ s are expected to be produced as secondaries in high-energy cosmic ray collisions, the energy of the antiprotons detected are too low to be accounted for by standard galactic cosmic ray models. Several possible explanations were discussed at the workshop: (1) The experiment is wrong. (2) The galactic models are grossly in error. (3) The $\bar{p}$ s were originally produced at high energies and then decelerated without annihilating. (4) Cosmic ray neutrons oscillate into antineutrons which then decay into antiprotons. (5) The antiprotons are primary cosmic rays originating from antigalaxies²⁵. Although there was no consensus on the interpretation of the experiment, there was agreement that if the experiment is correct, the implications will be very significant.

Galaxy formation

On the largest scales (>10 Mpc) the Universe is quite smooth and homogeneous, while on small scales (≤ 10 Mpc) it is lumpy (nuclei, atoms, molecules, creatures, planets, stars, galaxies, clusters and superclusters). The large-scale smoothness, the uniformity of the microwave background, and the absence of a profusion of black holes are evidence that as early as recombination ($T \sim 4,000$ K, $t \sim 10^6$ yr) the Universe could be described by an FRW model supplemented by small density perturbations. The concordance of the primordial nucleosynthesis calculations with the observed abundances of D and ${}^4\text{He}$, probably indicates that this description was valid as early as ~ 1 s after the 'bang' ($T \sim 10^{10}$ K). It is generally believed that small density fluctuations, present at recombination, after being freed from the pressure support provided by the radiation, collapsed due to their self-gravitational attraction and evolved into the structure we observe today. This is where the general consensus ends. The precise details of the growth of the density perturbations since recombination, their initial spectrum and origin are all areas of intense research and debate (for a review of galaxy formation, see ref. 26).

During the radiation dominated epoch ($T \geq 10^5$ K Ωh^2) density perturbations can be divided into two classes: isothermal and adiabatic. The subsequent development of structure is very different for these two types of initial fluctuations. Isothermal perturbations are characterized by uniform radiation temperature, with spatial variations in the baryon number-to-photon ratio. After recombination isothermal perturbations grow on all scales $\geq 10^5 M_\odot$ ($M_\odot = 1$ solar mass $\approx 2 \times 10^{33}$ g). Structure forms on the smallest scales first, which then themselves cluster,

and so on in a hierarchical manner. Adiabatic fluctuations are characterized by spatial variations in the temperature, but constant baryon number-to-photon ratio. During recombination, fluctuations on scales $\leq 10^{14} M_{\odot}$ are damped by photon diffusion, so that the first structures to form are $10^{14} M_{\odot}$ pancake-like objects (which are now identified with superclusters), which subsequently hydrodynamically fragment into galaxies, and so on.

Two recent developments in particle physics and cosmology have had important implications for theories of galaxy formation: baryosynthesis and the possibility of non-zero neutrino masses in the range 10–100 eV. Shortly after the idea of baryosynthesis was first discussed it was shown that if the Universe was nearly FRW during the epoch of baryosynthesis, it was not possible to produce a baryon asymmetry and isothermal fluctuations simultaneously—thus favouring the adiabatic picture²⁷. However, at the Santa Barbara workshop two groups reported that in a universe initially dominated by large-scale inhomogeneous shear, isothermal perturbations can be produced during baryosynthesis^{28,29}.

The adiabatic picture predicts temperature fluctuations of $0(10^{-3})$ in the microwave background on small angular scales ($\theta \geq 10''$), which do not agree with the observations. Massive neutrinos may save the adiabatic hypothesis. Last year a Russian group reported evidence that the electron neutrino has a mass $0(30 \text{ eV})$ (ref. 30) (GUTs can very naturally accommodate such neutrino masses). A neutrino of mass $\sim 3 \text{ eV}$ would dominate the mass density of the Universe; a neutrino of mass $\sim 100 \text{ eV}$ would provide closure density. Bond and Szalay reported that when massive neutrinos are added to the adiabatic picture, smaller initial perturbations are needed, resulting in predicted microwave fluctuations which are smaller than the present observational upper limits^{31,32}.

Both laboratory experiments (neutrino masses, proton decay, and so on) and astrophysical observations (more precise measurements of both the large and small scale structure of the 3 K background, further studies of the distribution of galaxies, and so on) should have a great impact on theories of galaxy formation in the next few years. During the workshop, it was clear that this was already beginning to occur. Silk discussed how the recently observed quadrupole anisotropy in the 3 K background tightly constrains both the adiabatic and isothermal pictures³³, and Davis reported on the recent observation of large holes in space (devoid of galaxies)³⁴. These holes are most easily explained in the adiabatic theory, while the observed quadrupole asymmetry favours the isothermal theory.

Exotic relics from the big bang

As the energies explored by manmade machines ($< 1,000 \text{ GeV}$) are quite modest in terms of the energy scales of GUTs ($\geq 10^{14} \text{ GeV}$), the early Universe is the only laboratory where physics at such energies can be studied. However, because this accelerator has not been in operation for $\sim 10,000 \text{ Myr}$, to study the physics of such high energies the observationalist must often search for fossil evidence although energies of up to 10^{19} GeV were probably realized in the early stages of the hot big bang. One example that received a lot of attention at Santa Barbara is the search for relic supermassive magnetic monopoles.

Polyakov and 't Hooft^{35,36} have shown that whenever a simple group (most of the gauge groups proposed for grand unification are simple groups) spontaneously breaks down to a group which contains a $U(1)$ factor [for example, to the low energy symmetry group, $SU(3) \times SU(2) \times U(1)$] magnetic monopoles with a mass of order 100 times the symmetry breaking scale exist. These monopoles are topological objects associated with 'kinks' in the Higgs field (the Higgs mechanism is used to break the symmetry spontaneously). It is difficult to see how less than ~ 1 monopole per horizon volume would be created, as causality precludes straightening out the Higgs field on larger scales. (The horizon volume is the largest-sized region that could have communicated by light signals since the 'bang'.) The usual symmetry-breaking scheme of GUTs results in a predicted relic

monopole abundance of about one per nucleon³⁷. As these monopoles have a mass of $\sim 10^{16} \text{ GeV}$ ($\sim 10^{-8} \text{ g}$), it is not surprising that such a present abundance can be ruled out by using our rather meagre knowledge of the present mass density of the Universe ($\rho \approx 4 \times 10^{-29} \text{ g cm}^{-3}$, which corresponds to $\Omega \approx 2$). A monopole abundance of one per nucleon would correspond to $\Omega \approx 10^{14}$.

This discrepancy between prediction and observation has been widely discussed. Possible solutions discussed at the workshop included mechanisms that would confine monopoles and antimonopoles in much the same way that quarks are confined in nucleons. In such schemes the confined monopoles and antimonopoles can annihilate to an acceptable level. Another possible solution is to keep the symmetry from being restored at high temperatures, as the monopoles are produced when the symmetry breaks. Finally, the most popular solution seems to be a hypothesis in which the universe supercools thereby delaying the GUT phase transition (for example, the inflationary universe of Guth³⁸). Delaying the phase transition reduces the number of monopoles produced, as the horizon volume grows with time and only about one relic monopole is produced per horizon volume. As no suppression mechanism yet proposed seems compelling, the number of monopoles present today, if any, is hard to determine. Clearly some observational information would be useful.

Past searches for magnetic monopoles have usually looked for the ionization caused by the passage of a monopole through matter. Because the monopoles of GUTs are expected to be heavy and slow, the past experimental limits do not apply. At Santa Barbara, Cline outlined his proposal to search for supermassive magnetic monopoles at a huge steel plant in Mt Iron, Minnesota. Cline believes that the relic monopoles that have bombarded the Earth since its formation became bound to ferromagnetic domains. Heating iron ore above the Curie point destroys the domains, thus releasing the monopoles. The perfect place to do the experiment is at the Mt Iron processing plant run by United States Steel, where there is a pipe which carries 18 million tons per year of iron-ore above the Curie point, releasing their pent-up monopoles to be detected as they fall in the Earth's gravitational field. Results of such searches will be awaited by both particle physicists and cosmologists. The discovery of a magnetic monopole would be direct evidence for grand unification and the spontaneous symmetry breaking of the GUT.

On a much smaller scale, but just as fundamental to both particle physics and cosmology, is the continued claim of the observation of fractional charge by a low temperature experimental group at Stanford headed by Fairbank³⁹. The importance of this experiment to particle physics is obvious, but cosmologists also have a stake in the results. Just as the monopoles were the fossil remains of the early Universe at temperatures of 10^{15} GeV , fractional charge may be the remnant of the period in the early Universe when hadrons were formed out of the quark gas.

The most obvious interpretation of the Fairbank experiment is the existence of free, fractionally-charged quarks. However, it is generally believed (but not rigorously proved) that in the $SU(3)$ colour gauge theory of quarks and gluons, coloured objects (for example, quarks) must always be bound into colourless states which have integral charge. For this reason the Stanford experiment has worried particle physics theorists. One possibility is that $SU(3)$ is also spontaneously broken, but at a very small energy scale, $\sim 10\text{--}50 \text{ MeV}$ (refs 40, 41). This allows for the existence of free coloured objects, like quarks. Other possibilities exist: the Stanford group could be observing leptons (which are not coloured) with non-integral charge, or perhaps they are seeing bound states of quarks and other integrally charged coloured objects which would result in fractionally-charged, colourless objects. Unfortunately, as particle physics does not supply a unique model to explain the existence of fractionally-charged states, cosmologists cannot give a unique answer for the expected abundance from production in the early

Universe. Wagoner reported that the relic abundances predicted from various models range from uncomfortably large to undetectably small⁴².

Phase transitions

During the 30 or so orders of magnitude of thermal history of the Universe ($T \sim 10^{32} \text{ K} \rightarrow 3 \text{ K}$) many phase transitions associated with spontaneous symmetry breaking (SSB) should have taken place⁴³, including the SSB of the GUT, of $SU(2) \times U(1)$, and perhaps the SSB of other intermediate symmetry stages. During these phase transitions it is likely that interesting phenomena occurred: during the GUT transition baryosynthesis and monopole production, and in all the stages of SSB, the possible production of entropy, creation of structure (that is, primordial density fluctuations), and generation of a cosmological term.

The standard mechanism (or crutch) for SSB is the Higgs mechanism. The details of the SSB transition depend on the parameters of the Higgs sector of the theory. If the transition is first order, there may be a large entropy release, which if it occurs after baryosynthesis dilutes the baryon asymmetry. (The baryon asymmetry is more correctly quantified as the baryon-to-entropy ratio rather than the baryon-to-photon ratio; however, today the two are essentially equivalent as most of the entropy in the Universe is in the 3 K background.) In the simplest model of SSB (Coleman–Weinberg symmetry breaking) it has been shown that the large entropy release associated with the SSB of $SU(2) \times U(1)$ ($T \sim 300 \text{ GeV}$) would dilute the baryon asymmetry produced during baryosynthesis by an unacceptably large factor, that is, leading to $B \leq 10^{-11}$ today^{44–46}.

More intriguing are the possible cosmological effects associated with the vacuum energies which result from SSB. The vacuum energy associated with the Higgs condensate corresponds to a uniform energy density which acts like a cosmological term (Λ), and which changes by $O(T_c^4)$ during a phase transition which occurs at temperature T_c . We know that the cosmological term today is not significant, $(\Lambda/8\pi G) \leq 4 \times 10^{-29} \text{ g cm}^{-3} \approx 10^{-46} \text{ GeV}^4$. It seems that unless one specifies an initial compensatory, 'true' cosmological term to cancel the 'induced' ones, the Universe would be left with an enormous cosmological constant due to SSB. Guth has discussed the so-called inflationary universe, in which the cosmological term temporarily dominates the dynamics of the Universe^{38,47}. An initial compensatory cosmological constant, $\Lambda_i/8\pi G \sim O(T_c^4)$, is initiated to cancel the induced cosmological term associated with the GUT phase transition ($T_c \sim 10^{15} \text{ GeV}$). If the phase transition proceeds slowly, then the cosmological term Λ_i will soon begin to dominate the dynamics and will do so until it is cancelled when the transition has been completed. During this period the Universe undergoes a de Sitter phase in which the expansion is exponential⁴⁷. The de Sitter phase has several potentially beneficial effects. First, the horizon (distance over which light signals can propagate) is inflated to encompass most of the observed Universe today, whereas in the usual FRW model, the horizon at the GUT epoch contains only 1 baryon. This could explain why on the largest scales the Universe is so smooth, as a patch of the Universe which could have been smoothed by particle interactions before the de Sitter phase is inflated to an enormous size during this epoch. The inflated horizon might also allow for the production of large-scale (galactic size and larger) density fluctuations which are necessary for galaxy formation. As mentioned earlier, the large horizon might possibly provide a solution to the monopole problem. When the transition has been completed, there is a large entropy release (baryosynthesis proceeds after this). This could explain the Universe has expanded so long without recollapsing or becoming negative—curvature dominated (in units of the fundamental time unit, $t_{\text{Planck}} \sim 10^{-43} \text{ s}$, $\tau_{\text{Universe}} \sim 10^{60} t_{\text{Planck}}$). The curvature term grows in importance with time; as it is not significant today, earlier it must have been far less significant than the energy density term. Why this is the case is a mystery (the so-called 'flatness problem'). However, the tremendous entropy release would provide a natural explanation because it

allows the energy density to increase relative to the curvature term by a very large factor. Guth's 'inflationary scenario' is not without its complications. First, there is the problem of gracefully returning to FRW behaviour; and second, it presupposes the reality of these enormous induced cosmological terms associated with SSB. The cosmological term today is essentially zero; perhaps we should adopt the point of view that it has always been zero (for reasons we do not understand).

Fundamental problems

The 6-month program and week-long intensive workshop at Santa Barbara were both evidence that cosmology and particle physics is a young field that has come of age. As we have tried to stress here, the symbiotic relationship which exists between these two fields is one in which information flows in both directions: cosmology can constrain particle physics theories and particle physics can have profound implications for cosmology. Cosmological constraints restrict the class of viable theories, rather than provide guidance or direction. However, there is the hope that particle physics theories, GUTs in particular, might help to resolve some of the fundamental problems of classical (that is, non-quantum) cosmology. To conclude, we shall try to assess the impact to date and the possible future impact of GUTs on what we consider to be the six fundamental puzzles of classical cosmology⁴⁸.

The most significant contribution of grand unification is baryosynthesis. Although we are still a long way from being able to predict the precise magnitude, or even the sign of the baryon asymmetry with GUTs, for the first time a framework exists for understanding why there is a baryon asymmetry when the laws of physics at the microscopic level are so nearly symmetrical.

The introduction of GUTs to cosmology, while appearing to have resolved one puzzle, unfortunately seems to have created another: the monopole problem. All the straightforward GUT scenarios of the early Universe predict that relic, superheavy monopoles ($M \geq 10^{16} \text{ GeV}$) should today provide enough mass density to close the Universe many times over ($\Omega \geq 10^{14}$). Perhaps, the solution is already at hand; however, it is equally likely that the resolution of the problem will be more subtle, and may even come at a very fundamental level with other important implications. The detection of relic monopoles and the determination of their cosmic abundance would be of the greatest significance for both particle physics and cosmology.

Another fundamental problem is the 'flatness' problem⁴⁹ which can also be regarded as an 'age' problem. There is only one fundamental time scale: $t_{\text{Planck}} \sim 10^{-43} \text{ s}$; the age of the Universe is $\sim 10^{60} t_{\text{Planck}}$. In newtonian language, the 'flatness' problem is: the initial (at t_{Planck}) kinetic and potential energies of the Universe must have been equal (and opposite) to an accuracy of some 50 or so decimal places, otherwise the Universe would have long since recollapsed or entered a coasting phase (dynamics dominated by negative curvature). Guth's 'inflationary universe' may answer this problem, but at present there are other serious problems inherent to the scenario.

Next, there is the homogeneity problem: on the largest scales ($\geq 10 \text{ Mpc}$) the Universe is very smooth, while on smaller scales it is quite lumpy (stars, galaxies, clusters, and so on). In the standard FRW cosmology, the large-scale homogeneity cannot be explained by causal processes, because causally distinct regions are uniform to a high degree of accuracy. Regions of the sky separated by more than $\sim 1^\circ$ or so were causally disconnected when the microwave photons we detect today were emitted, and yet have temperatures which are equal to within ~ 1 part in 10^3 . Given the 'proper spectrum' of initial small inhomogeneities (which is also a matter of disagreement), the present structure can be understood as being a result of the growth of the small fluctuations due to gravitational instability. The origin of the initial inhomogeneities is still a mystery which is also exacerbated by the existence of particle horizons in FRW models. The creation of fluctuations at very early times (for example, during a phase transition) is severely limited by the size of the horizon, and for this reason the resulting spectrum of

perturbations is usually far too small to be of any importance. Again, Guth's scenario may help to resolve both parts of the homogeneity problem, as the horizon problem is effectively circumvented.

GUTs thus far have shed no light on why the present cosmological term is so small. As is well known, quantum field theories (for example, quantum electrodynamics) predict an infinite energy density associated with the vacuum, which results in an infinite cosmological constant. (The Casimir effect strikingly demonstrates the reality of the vacuum energy density.) With a spontaneously broken gauge theory the situation is even more perplexing as the vacuum energy density changes during each stage of SSB. Even if one chooses to ignore the infinite contribution to the cosmological term, one must face the induced cosmological constants associated with each stage of SSB. If an initial compensatory, 'true' cosmological term is introduced to cancel the induced ones, then to cancel the cosmological constant which results from the SSB of the GUT, the initial cosmological constant must be specified to an accuracy of >100 decimal places. It is the assumption of the reality of these induced cosmological terms associated with SSB that leads to Guth's 'inflationary universe'. It is perhaps equally reasonable to suppose that as the cosmological term is essentially zero today, it has always been zero (for an unknown reason). It seems clear that a satisfactory solution to the cosmological constant problem will involve connections between cosmology and quantum field theory at a very deep level.

GUTs have also had very limited impact on the isotropy problem: why is the Universe isotropic to such a high degree of precision? The isotropy is evidenced by the 3 K background, galaxy counts, and the X-ray background.

The solution to some or even all of the five unresolved conundrums may well involve quantum gravity. As was discussed by Hartle, Hu and Parker at the Santa Barbara workshop, there are indications that particle creation due to quantum gravitational effects is highly efficient in damping initial anisotropy⁵⁰, and that quantum gravitational effects may remove particle horizons⁵¹⁻⁵⁴ from cosmological models, perhaps leading to a solution to both parts of the horizon problem. However, given the magnitude and fundamental nature of the problem of formulating a quantum theory of gravity, it seems most prudent not to pin one's hopes on quantum gravity, but to pursue solutions that do not involve quantum gravity. It is not implausible that GUTs could provide solutions to most of these five puzzles.

We benefited greatly from conversations with members of the Early Universe Program and participants at the week-long workshop in Santa Barbara at the NSF's ITP. We especially thank the organizers of the Early Universe Program (G. Steigman and J. Hartle), of the Cosmology and Particle Physics Workshop (D. N. Schramm, G. Steigman and J. Hartle), and the director of the ITP, Walter Kohn. This work was supported by the NSF PHY 77-27084 at Santa Barbara and by the DOE at Los Alamos and Chicago (DE-AC02-80ER10773).

1. Weinberg, S. *Gravitation and Cosmology* Ch. 15 (Wiley, New York, 1972).
2. Misner, C., Thorne, K. & Wheeler, J. *Gravitation* Ch. 27-30 (Freeman, San Francisco, 1973).
3. Landau, L. & Lifshitz, E. *The Classical Theory of Fields* 4th edn, Ch. 14 (Pergamon, Oxford, 1975).
4. Georgi, H. *Scient. Am.* **244** (4), 40 (1981).
5. Goldman, T. & Ross, D. *Nucl. Phys.* **B162**, 102 (1980).
6. Weinberg, S. *Scient. Am.* **244**(6), 64 (1981).
7. Miyake, S. *2nd Workshop on Grand Unification* (Ann Arbor, 1981).
8. Schramm, D. & Wagoner, R. A. *Rev. Nucl. Sci.* **27**, 37 (1977).
9. Olive, K., Schramm, D., Steigman, G., Turner, M. & Yang, J. *Astrophys. J.* **246**, 557 (1981).
10. Yang, J., Turner, M., Steigman, G., Schramm, D. & Olive, K. (in preparation).
11. Stecker, F. *Phys. Rev. Lett.* **44**, 1237, 1437(E) (1980).
12. Rees, M. *Nature* **275**, 35 (1978).
13. Carr, B. & Rees, M. *Astr. Astrophys.* **61**, 705 (1977).
14. David, Y. & Reeves, H. in *Proc. Les Houches Summer School* Vol. 32 (eds Balain, R., Audoze, J. & Schramm, D.) (North Holland, New York, 1980).
15. Harvey, J. & Kolb, E. *Phys. Rev. D* (in the press).
16. Steigman, G. A. *Rev. Astr. Astrophys.* **14**, 339 (1976).
17. Wilczek, F. *Scient. Am.* **243**(6), 60 (1980).
18. Waldrop, M. *Science* **211**, 803 (1981).
19. Sakharov, A. *Zh. ETF Pis'ma* **5**, 32 (1967).
20. Weinberg, S. *Phys. Rev. Lett.* **42**, 850 (1979).
21. Toussaint, D. et al. *Phys. Rev. D* **19**, 1036 (1979).
22. Turner, M. *Baryosynthesis Without Out-of-equilibrium Decay* (UCSB preprint NSF-ITP-81-30).
23. Ellis, J., Gaillard, M., Nanopoulos, D. & Rudaz, S. *Phys. Lett.* **99B**, 101 (1981).
24. Buffington, A. et al. *Astrophys. J. Lett.* (in the press).
25. Stecker, F., Protheroe, R. & Kazanas, D. *Phys. Rev. Lett.* (submitted).
26. Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, New Jersey, 1980).
27. Turner, M. & Schramm, D. *Nature* **279**, 303 (1979).
28. Barrow, J. & Turner, M. *Nature* **291**, 469 (1981).
29. Bond, J., Kolb, E. & Silk, J. *Astrophys. J.* (in the press).
30. Lyubimov, V. et al. *Phys. Lett.* **94B**, 266 (1980).
31. Bond, J. & Szalay, A. (in preparation).
32. Doroshkevich, A. et al. *Proc. 10th Texas Symp.* (1981).
33. Silk, J. & Wilson, M. *Astrophys. J. Lett.* **244**, L37 (1981).
34. Davis, M. & Press, W. (in preparation).
35. 't Hooft, G. *Nucl. Phys.* **B79**, 276 (1974).
36. Polyakov, A. *JETP Lett.* **20**, 194 (1974).
37. Preskill, J. *Phys. Rev. Lett.* **43**, 1365 (1979).
38. Guth, A. *Phys. Rev. D* **23**, 347 (1981).
39. La Rue, G., Phillips, J. & Fairbank, W. *Phys. Rev. Lett.* **46**, 967 (1981).
40. De Rujula, A., Giles, R. & Jaffe, R. *Phys. Rev. D* **17**, 285 (1978).
41. Slansky, R., Goldman, T. & Shaw, G. *Phys. Rev. Lett.* (in the press).
42. Wagoner, R. & Steigman, G. *Phys. Rev. D* **20**, 825 (1979).
43. Linde, A. *JETP Lett.* **19**, 183 (1974).
44. Sher, M. *Phys. Rev. D* **22**, 1989 (1980).
45. Guth, A. & Weinberg, E. *Phys. Rev. Lett.* **45**, 1131 (1980).
46. Witten, E. *Nucl. Phys.* **B177**, 477 (1981).
47. Kolb, E. & Wolfman, S. *Astrophys. J.* **229**, 428 (1980).
48. Turner, M. in *Proc. 2nd Workshop on Grand Unification* (1981).
49. Dicke, R. & Peebles, P. J. E. in *General Relativity: An Einstein Centenary Survey* (eds Hawking, S. & Israel, M.) (Cambridge University Press, 1979).
50. Hu, B. *Phys. Lett. B* (in the press).
51. Parker, L. & Fulling, S. *Phys. Rev. D* **7**, 2357 (1973).
52. Fischetti, M., Hartle, J. & Hu, B. *Phys. Rev. D* **20**, 1757 (1979).
53. Starobinsky, A. *Phys. Lett.* **91B**, 99 (1980).
54. Zee, A. *Phys. Rev. Lett.* **44**, 703 (1980).

ARTICLES

No evidence of rings around Neptune

J. L. Elliot*, D. J. Mink*, J. H. Elias†, R. L. Baron*, E. Dunham*, J. E. Pingree*, R. G. French*, W. Liller‡, P. D. Nicholson§, T. J. Jones§ & O. G. Franz||

* Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Cerro Tololo InterAmerican Observatory, Casilla 63-D, La Serena, Chile

‡ Center for Astrophysics, Cambridge, Massachusetts 02138, USA

§ Mount Stromlo and Siding Spring Observatories, The Australian National University, Canberra, ACT 2600, Australia

|| Lowell Observatory, Flagstaff, Arizona 86002, USA

Observations of two stellar occultations by Neptune were used to search for possible equatorial and polar rings. No ring occultation events were identified, and an upper limit of 0.07 can be placed on the optical depth of any equatorial rings greater than 5 km wide with radii greater than 31,400 km. Any ring system of Neptune must be much less extensive than the ring systems of Uranus and Saturn, but a jovian-type ring of low optical depth would have escaped detection by our search.

THE unexpected discovery¹ of rings around Uranus in 1977 excited new interest in planetary rings and the extreme narrowness of its nine rings—some only a few kilometres wide in the radial direction—created a new puzzle for dynamicists^{2,3}. The interest increased in 1979, when a ring system was found around Jupiter⁴. The three known ring systems appeared strikingly different: Jupiter's, of extremely low optical depth; Saturn's, broad and bright; and the uranian rings, dark and narrow. Voyager 1 established an important similarity between the saturnian and uranian rings, when its high resolution images revealed that a multitude of narrow saturnian rings constitute the several broad rings that we see from Earth⁵.

Rings around Neptune?

Reports of visual detection of rings around Neptune, made shortly after the planet's discovery in 1846, have been either withdrawn or unconfirmed⁶. Furthermore, the question of whether Neptune has rings cannot be answered on theoretical grounds, as the origins of the three known ring systems have not been established and may all be different⁷⁻¹². One might suspect that because three of four giant planets have rings, Neptune should also—assuming a ring system results directly or indirectly from the formation of a giant planet. On the other hand, the three planets with rings also have regular satellite systems, but Neptune's satellite system is peculiar¹³. It has only two confirmed satellites: Triton, a large satellite with an inclined retrograde orbit, and Nereid, a distant satellite with a highly eccentric orbit.

Techniques

The most sensitive techniques for detecting planetary rings have been Voyager imagery and stellar occultations¹⁴. Since Voyager 2 is not due to explore the Neptune system until 1989, we have selected from lists of possible occultations by the Neptune the most appropriate events for detecting a ring system. Using occultations to search for rings around Neptune is more difficult than using these to observe the uranian rings, because an equatorial ring system around Neptune would subtend an angle of only 2 arcs perpendicular to the motion of Neptune; the uranian rings subtend an angle of 8 arcs perpendicular to the motion of Uranus. Furthermore, a complete search for rings around Neptune must include the possibility of rings in the non-equatorial, yet stable, orbits described by Dobrovolskis¹⁵. The only previously reported occultation observations are those of Nicholson and Jones¹⁶, who obtained incomplete data for an occultation on 21 August 1980. They describe a possible ring event (at a projected equatorial radius of 1.5 R_N), but there have been no corroborating data.

The two occultations used in our search were (1) star no. 28 in the list of Mink *et al.*¹⁷ which occurred on 10 May 1981, and (2) an uncatalogued star that we identified as an observable event only a few days before it occurred on 24 May 1981. As shown in Table 1, the first event was observed from various sites around the Pacific Ocean, while the second was observed only from Cerro Tololo. Observations were made at several wavelengths

Table 1 Ring search observations

Date	Observatory	Telescope† aperture (m)	Filter*	Observers
10 May 1981	Mauna Kea	3.0	K	Elliot, Mink
		2.2	B, R ₁	Liller, Pingree
		0.6	R ₂	Franz, French
10 May 1981	Siding Spring	1.0	B, R ₁	Dunham, Zarro
10 May 1981	Mount Stromlo	1.9	K	Jones, Nicholson
		0.8	B, R ₁	Baron, Smith
24 May 1981	Cerro Tololo	4.0	K	Elias, Martin

* The filters had the following centre wavelengths and passbands (FWHM): B, 0.44 and 0.10 μm ; R₁, 0.78 and 0.18 μm ; R₂, 0.87 and 0.06 μm ; K, 2.2 and 0.4 μm .

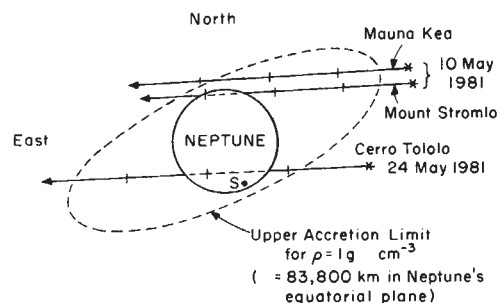


Fig. 1 Regions probed for rings around Neptune. The solid lines show the tracks of the occulted stars relative to Neptune as seen from Mauna Kea, Mount Stromlo and Cerro Tololo. The tracks begin at 11:00 UTC for the 10 May event and at 2:00 UTC for the 24 May event. The tic marks on each track occur at 30-min intervals. The dashed ellipse shows the upper accretion limit²² in Neptune's equatorial plane. Inside this limit, the tidal force from Neptune exceeds the gravitational attraction for spherical particles of equal size. Here, such particles could not accrete into larger bodies, but would remain as rings (see note added in proof).

with equipment described elsewhere¹⁸⁻²⁰. In each case the observers recorded (on magnetic tape) the intensity of starlight as the star passed behind the region around Neptune that could accommodate rings.

Results

Figure 1 shows a plot of the tracks of the stars as they would appear in the sky, the dashed portions indicating where Neptune occulted the star. The duration of the Neptune occultations, along with the occultation radius of Neptune obtained from Taylor's circular solution²¹, were used to establish the positions of the tracks with respect to Neptune. Also shown is the upper accretion limit for material of $\rho = 1 \text{ g cm}^{-3}$ in the equatorial plane of Neptune. Within the upper accretion limit, the gravitational attraction of spherical bodies of equal size is less than the tidal force from Neptune²². Hence, spherical particles of equal size would remain as rings, and not condense into satellites, inside this limit. For particles of unequal size, the accretion limit would be closer to Neptune. The direction of Neptune's pole in Fig. 1 was assumed to coincide with the pole about which Triton's orbit precesses^{23,24}. Our assumed direction could differ by a few degrees from the actual direction of Neptune's pole, which cannot be calculated because Triton's mass is uncertain. The sky-plane velocity of the star was 19 km s^{-1} for the first event and 22 km s^{-1} for the second.

The two events combined cover the space for rings around Neptune. Fortunately for this search, Neptune did not occult the star as seen from Mauna Kea, so that this track covers the

Table 2 Regions probed for rings

UTC on 10 May 1981	Projected equatorial distance (km)			UTC on 24 May 1981	Projected equatorial distance (km) Cerro Tololo
	Mauna Kea	Siding Spring	Mount Stromlo		
10:00	182,200	190,900	190,900	2:00	130,900
10:30	138,300	146,000	146,000	2:30	77,300
11:00	97,900	102,800	102,700	3:00	34,300
11:30	67,900	65,100	64,700	3:30	52,800
12:00	64,700	48,500	47,800	4:00	104,100
12:30	91,300	70,500	69,700		
13:00	130,400	109,900	109,300		
13:30	173,800	154,100	153,400		

Projected equatorial distances for times other than those tabulated should be obtained from three-point interpolation of the tabulated values. Figure 1 can be used to determine when the star was behind the planet and we could not probe for rings; Fig. 2 shows the time intervals when data were being acquired. Minimum equatorial distances are: MKO, 61,700 km at 11:48 UTC; SSO, 48,400 km at 11:58 UTC; MSO, 47,600 km at 11:58 UTC; and CTIO, 31,400 km at 3:08 UTC.

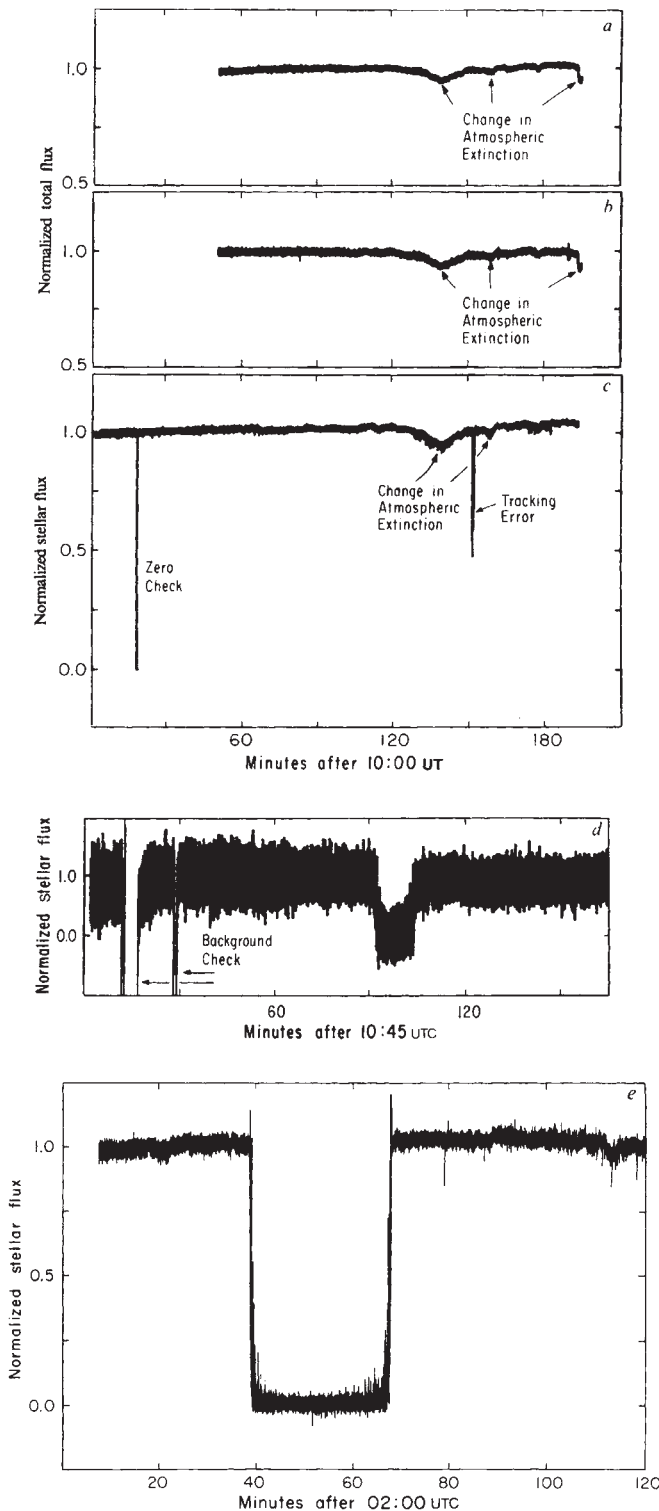


Fig. 2 Intensity versus time for: *a-c*, the 10 May event from Mauna Kea; *d*, the 10 May event from Mount Stromlo; *e*, the 24 May event from Cerro Torolo. All data have been averaged at 0.2s. Frames *a*, *b* show the combined light from Neptune and the star observed simultaneously at 0.44 μm and 0.78 μm with the 2.24-m telescope. Frame *c* shows the intensity at 2.2 μm observed with the IRTF. All notable dips in signal can be ascribed to causes other than ring occultations. In *d* the times of immersion and emersion of the star behind Neptune were used to calculate the apparent position of the star relative to Neptune. In *e* the large increases in signal at immersion and emersion by Neptune are the occultation spikes, caused by density variations in the atmosphere by Neptune. The dip in signal at 3:18:41 UTC is too narrow to have been caused by a ring occultation. The remaining conspicuous dips could not have been caused by continuous equatorial rings, or corresponding dips would have been observed during the 10 May occultation.

possibility of polar rings. For equatorial rings, Table 2 gives the projected distance from the centre of Neptune in its equatorial plane, as a function of time, for all four stations. The minimum distance probed was from Cerro Tololo, just as the star emerged from behind the planet (at a point corresponding to 31,400 km [$1.3 R_N$] in the equatorial plane).

The observed intensities versus time are shown in Fig. 2 for Mauna Kea, Mount Stromlo and Cerro Tololo. The averaging time was chosen to match the width of a diffraction-limited occultation profile, which would have been produced by a ring narrower than ~ 4.6 km (refs 25, 26). Note that the star occulted on 10 May had a projected diameter at Neptune of ~ 10 km (inferred from *JHK* photometry), which would broaden any ring occultation profiles to about twice their diffraction limit.

First we examine the data obtained with the 2.24-m telescope and IRTF (NASA's 3.0-m IR telescope facility), displayed in Fig. 2 *a-c*. As the spectra of Neptune and the star are not similar, the relative contributions of each are quite different at the three wavelengths observed. For the blue channel, the contribution of the star is $< 1\%$ of the total signal; for the red channel, the star contributes 21% and for the 2.2- μm channel the stellar contribution is $> 99\%$. Hence, a partial occultation of the star by material near Neptune would not be detectable in the blue channel, and would have different fractional drops in the red and 2.2- μm channels. On the other hand, a change in atmospheric extinction (assuming it was not wavelength dependent) would cause approximately the same fractional drop in total flux for all three channels. A third possible cause of a drop in signal, a guiding or tracking error by one of the telescopes, would appear in the data from one telescope.

All observed drops in signal have explanations other than occultations by possible rings. These have been marked in Fig. 2. The small dip in the red channel near 11:20 UTC occurred during a focusing adjustment of the telescope. The remaining dips in the IRTF signal are $< 5\%$ of the stellar intensity. Hence, we conclude that any rings of Neptune within the region probed from Mauna Kea, (> 5 km wide in the sky plane plane projection) must have an optical depth < 0.05 .

The data from the 0.8-m telescope at Mount Stromlo, shown in Fig. 2*d*, contain the occultation that we used to calculate the path of the star with respect to Neptune. Although the data are noisier than those obtained from Mauna Kea, the star as seen from Mount Stromlo probed closer to Neptune, and these data would have revealed any rings similar to the α , β , γ , δ and ϵ rings of Uranus¹⁴. The 2.2- μm data from the 1.9-m telescope at Mount Stromlo (not displayed) have much better signal-to-noise ratios than those obtained with the 0.8-m telescope, and nine dips in signal occurred in these data that were candidates for ring occultations. However, all these events happened at times when corresponding events would have been visible from Mauna Kea as well, if they were caused by rings. The lack of corresponding events at Mauna Kea leads us to conclude that the events recorded with 1.9-m telescope were caused by thin clouds and/or poor seeing. If any had been due to a total occultation by a small satellite, a corresponding event would have been recorded with the 0.8-m telescope. No total occultations, except for Neptune, were recorded (see Fig. 2*d*).

The data from the occultation by Neptune on 24 May 1981 are shown in Fig. 2*e*. The increases in signal above full-scale at planetary immersion and emersion are the spikes caused by density variations in Neptune's atmosphere. As this event was observed with only one telescope, and at a single wavelength, we had more difficulty in establishing the cause of dips in signal. The most conspicuous dip occurred at 03:18:41 UTC, which corresponds to an equatorial radius of $1.55 R_N$. Curiously, this is near the same equatorial radius of the possible ring occultation reported by Nicholson and Jones¹⁶. However, we believe both events had causes other than rings and their coincidence is fortuitous. At high resolution the CTIO event has a FWHM of only 2.7 km. This is about half the minimum width (4.6 km, due to diffraction^{25,26}) that could have been produced by a ring or small body at the distance of Neptune. Hence we conclude that

this dip was not caused by a ring or small body near Neptune. The other three notable dips in signal that occur after 3:25 UTC were apparently not caused by continuous equatorial rings either, as this region was probed during the 10 May event, with negative results. In fact, the dip at 3:25 UTC occurred when the planet was near the edge of the photometer beam. Other than the dips just discussed and the occultation by Neptune, the signal never drops below 0.93 of its unocculted value. Hence, we place a limit of 0.07 on the optical depth of any equatorial rings of Neptune, with widths > 5 km and radii $> 31,400$ km.

Our upper limit is about four times lower than would have been necessary to detect the least opaque uranian ring (ring 6), but we could not have detected a jovian-type ring, of low optical depth. (The jovian ring has not yet been detected by radio or optical occultation^{27,28}.) Hence, in terms of detectability by occultations, any system of rings around Neptune must be much less extensive than the uranian system.

In terms of the amount of material that could exist as neptunian rings, our limit on the optical depth allows a large range of masses. On one hand, we could have detected as little as 10^{12} g of material concentrated into a narrow ring. However, the form of the material least amenable to detection by our observations would be a collection of small satellites, ~ 1 km in diameter. This diameter would be too small for an individual satellite to cause an observable occultation. If the mean optical depth of the collection of satellites were < 0.07 , we would not have detected the ensemble either. If the material were carbonaceous ($\rho = 3$ g cm⁻³), the combined mass of these satellites could be as great as 4×10^{23} g. This is comparable to recent upper limits on the mass of Saturn's rings²⁹ and would equal that of a single satellite, 600 km in diameter.

Received 31 August; accepted 25 September 1981.

1. Elliot, J. L. *A. Rev. Astr. Astrophys.* **17**, 445 (1979).
2. Goldreich, P. & Tremaine, S. *Nature* **277**, 97 (1979).
3. Dermott, S. F., Murray, C. D. & Sinclair, A. T. *Nature* **284**, 309 (1980).
4. Smith, B. A. *et al. Science* **206**, 927 (1979).
5. Smith, B. A. *et al. Science* **212**, 163 (1981).
6. Baum, R. *The Planets, Some Myths and Realities*, 120 (David and Charles, Newton Abbot, 1973).
7. Smoluchowski, R. *Commun. Astrophys.* **8**, 69 (1979).
8. Pollack, J. B. *Space Sci. Rev.* **18**, 3 (1975).
9. Ip, W.-H. *Space Sci. Rev.* **26**, 39 (1980).
10. Ip, W.-H. *Space Sci. Rev.* **26**, 97 (1980).
11. Dermott, S. F. *Phil. Trans R. Soc.* (in the press).
12. Prentice, A. J. R. & ter Haar, D. *Moon Planets* **21**, 43 (1979).
13. Morrison, D., Cruikshank, D. P. & Burns, J. A. in *Planetary Satellites* (ed. J. A. Burns) 3 (University of Arizona Press, 1977).
14. Elliot, J. L. in *Uranus and the Outer Planets* (ed. G. A. Hunt) (Cambridge University Press, in the press).

Conclusions

It is puzzling to reconcile our failure to detect rings with the recent discovery (through occultation observations) of an apparently small, third satellite of Neptune³⁰, with an orbital radius of $3 R_N$. As the odds against discovering a lone body in this manner would be so great, one might conclude that Neptune has several, if not many, small satellites. One small satellite orbits just inside, and another just outside, the F ring of Saturn; at least two exist near Jupiter's ring and several small satellites have been postulated to be intimately associated with the uranian rings.

Hence we must still entertain the possibility of a neptunian ring system that we were unable to detect. A further opportunity to search for rings with occultations, with better sensitivity, occurs¹⁷ in 1983, and Voyager encounters Neptune in 1989. If these opportunities fail to reveal rings, then we need wait only $\sim 10^8$ yr until tidal interaction between Neptune and Triton will have caused Triton to move close enough to Neptune to break up into what should be a spectacular system of rings³¹.

We thank the telescope allocation committees of Mauna Kea, the IRTF, Mount Stromlo and Siding Spring for our observing time and B. Schaefer, F. Cheigh, G. Martin, R. Smith and D. Zarro for observational assistance. This work was partially supported by grants from NSF (AST-7827879 and AST-8010699) and NASA (NSG-7526 and NSG-7603).

Note added in proof: An error of $2\frac{1}{2}$ was found by P.D.N. in Smoluchowski's expression for the upper accretion limit²². For $\rho = 1$ gm cm⁻³, the correct radius of the upper accretion limit in Fig. 1 is 66,500 km (not 83,800 km).

15. Dobrovolskis, A. R. *Icarus* **43**, 222 (1980).
16. Nicholson, P. D. & Jones, T. J. *IAU Circ. No.* 3515 (1980).
17. Mink, D. J., Klemola, A. R. & Elliot, J. L. *Astr. J.* **86**, 135 (1981).
18. Frogel, Jay A. & Elias, J. H. *CITF Facilities Manual* Revised edn (1980).
19. Elliot, J. L., Veverka, J. & Goguen, J. *Icarus* **26**, 387-407 (1975).
20. Baron, R. L. *et al.* (in preparation).
21. Taylor, G. E. *Mon. Not. R. astr. Soc.* **147**, 27 (1970).
22. Smoluchowski, R. *Nature* **280**, 377 (1979).
23. Gill, J. R. & Gault, B. L. *Astr. J.* **73**, S95(A) (1968).
24. Harris, A. W. *Bull. Am. astr. Soc.* **12**, 705 (1980).
25. Elliot, J. L. *et al. Astr. J.* **86**, 127 (1981).
26. Nicholson, P. D., Matthews, K. & Goldreich, P. *Astr. J.* (in the press).
27. Tyler, G. L., Marouf, E. A. & Wood, G. E. *J. geophys. Res.* (in the press).
28. Dunham, E., Elliot, J. L., Mink, D. J. & Klemola, A. R. *Astr. J.* (submitted).
29. Null, G. W., Lau, E. L., Biller, E. D. & Anderson, J. D. *Astr. J.* **86**, 456 (1981).
30. Reitsema, H. J., Hubbard, W. B., Lebofsky, L. A. & Tholen, D. J. *Science* (in the press).
31. McCord, T. B. *Astr. J.* **71**, 585 (1966).

Global satellite measurements of water vapour, wind speed and wave height

Dudley B. Chelton, Kevin J. Hussey & Michael E. Parke

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

Data from the 100 days of Seasat observations in 1978 provided the first global maps of mean wind speed and wave height measured from satellites. They reveal previously unknown features in both fields and demonstrate the potential for satellite monitoring and forecasting of the worldwide sea state.

ON 28 June 1978 the United States launched Seasat, an experimental satellite to demonstrate the utility of microwave remote sensing of the ocean surface. On board were three active microwave radars (an altimeter, a scatterometer and a synthetic aperture radar) and two passive radiometers (a scanning multichannel microwave radiometer and a visible and IR radiometer). Because of their relative insensitivity to cloud cover compared with visual and IR sensors, microwave sensors have the advantage of all weather operations. A more detailed description of the Seasat mission is given in ref. 1. Although a short circuit in its electrical system ended the observations

prematurely on 10 October 1978, Seasat provided the first global pictures from space of wave height and wind speed. Seasat also measured atmospheric liquid water, water vapour, sea-surface temperature and sea-surface topography with better accuracy than from previous satellites.

Measurements

Results of global measurements of atmospheric water vapour by the Seasat Scanning Multichannel Microwave Radiometer (SMMR) and wave height and wind speed by the Seasat altimeter (ALT) are presented here. The primary objective of ALT

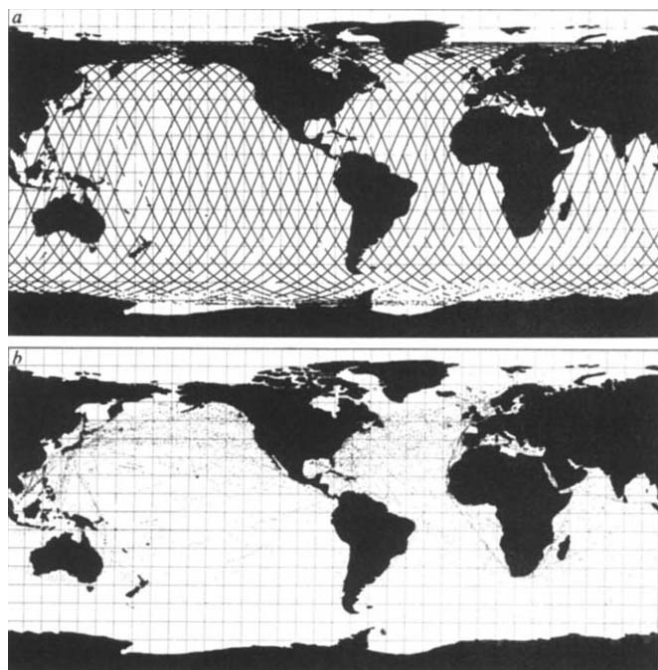


Fig. 1 Representative 3-day global data coverage by: *a*, the Seasat altimeter; and *b*, conventional ship weather reports (courtesy of D. McLain, National Marine Fisheries Service, Monterey, California). Both the examples shown are for 9–11 August 1978.

was to measure the height of the sea surface relative to a reference ellipsoid. However, the shape and power of the returned radar pulse are also used to infer wave height and wind speed at the sea surface. The 13.5-GHz Seasat altimeter is a successor of the GEOS 3 altimeter with a shorter pulse width (3.125 ns) and higher repetition rate (1,020 Hz) which lead to more precise measurements of sea surface height (better than 10 cm). The 1.59° half power beam width directed at the spacecraft nadir gives a cross track sea-surface footprint that increases from 2.4 km for 0 m wave height to 11.6 km for 20 m wave height. For the 1-s averages presented here, the along track sea-surface footprint ranges from 9.4 to 18.7 km. For typical wave heights of 3 m the ALT footprint is $\sim 5 \times 12$ km.

ALT operated intermittently, producing data over a total of 78 days. Because Seasat circled the Earth about 14 times each day at an altitude of ~ 800 km and an inclination of 108° , successive equatorial crossings moved in a westward direction with a separation of $\sim 2,500$ km. In 3 days, the equatorial spacing of the ground tracks was about 900 km. The full Seasat mission resulted in global coverage with ~ 135 km equatorial spacing of ground tracks. Typical 3-day ALT data are shown in Fig. 1*a*. The gaps in the data are due either to ALT being temporarily turned off or to a flagged error condition (such as anomalously high wave height or wind speed estimates or a satellite tilt angle beyond 0.5°). For comparison, global data from conventional ship weather reports are shown in Fig. 1*b* for the same 3-day period: ALT provides more than an order of magnitude increase in the quantity of data (131,700 compared with 7,600 observations). Even more importantly, the spatial coverage by ALT is more nearly uniform over the globe than the ship reports which tend to be concentrated along standard shipping routes in the Northern Hemisphere.

A detailed description of how the wave height can be extracted from the ALT radar return pulse is given in refs 2, 3. The waves stretch the leading edge of the returned radar pulse because of early returns from wave crests and later returns from wave troughs. Thus, the slope of the leading edge of the returned radar pulse is inversely related to the wave height. The shape of the returned pulse has been empirically related to the significant wave height (SWH) defined to be four times the r.m.s. of the crest-to-trough wave height, roughly equivalent to the average

height of the one-third largest waves present in the ALT footprint. Computations of wave heights were carried out on board the satellite providing real-time wave measurements. Two simple bias corrections were later applied on the ground: a pre-launch calibration bias removal and a bias correction due to satellite tilt angle (flagged if $>0.5^\circ$). A scatter plot comparing 51 buoy measurements with the ground-processed ALT estimates of SWH is shown in Fig. 2*a* (see ref. 3). The mean and r.m.s. differences (ALT minus buoy) over this sample data set are small (28 and 38 cm respectively). Apparently, the on-board processor estimates of SWH > 2 m may be biased ~ 50 cm high but the present sample size is too small to confirm this.

The altimeter antenna measures specular return from the nadir sea surface which is primarily determined by that part of the ocean wave spectrum with wavelengths longer than the ~ 2 -cm wavelength of the incident radiation. As the wind speed increases, the sea-surface roughness increases and a greater fraction of the incident radiation is scattered away from the satellite. Thus the power of the returned radar pulse is inversely related to the wind speed. A more detailed description of the relationship between radar backscatter and the wind speed is given in ref. 4. A scatter plot comparing 87 buoy measurements with ALT estimates of wind speed is shown in Fig. 2*b* (see ref. 3). A direct comparison of Seasat and GEOS 3 altimeter measurements during the overlap of the two satellite missions indicated that the Seasat backscatter values were biased 1.6 dB higher than GEOS 3 (see ref. 3). The present wind estimates were generated from the Seasat ALT backscatter values by first removing this 1.6-dB bias and then using a wind speed algorithm previously derived for the GEOS 3 altimeter (see ref. 3). The mean and r.m.s. differences (ALT minus buoy) from this limited sample size are 0.25 m s^{-1} and 1.58 m s^{-1} respectively. The lack of data means that the reliability of the relationship at wind speeds in excess of 10 m s^{-1} is questionable. However, experience with GEOS 3 indicates that an altimeter can provide reliable wind speed estimates to at least 20 m s^{-1} . Investigations are underway to determine the accuracy of the Seasat ALT wind speed estimates over a broader range of conditions and from a larger data base.

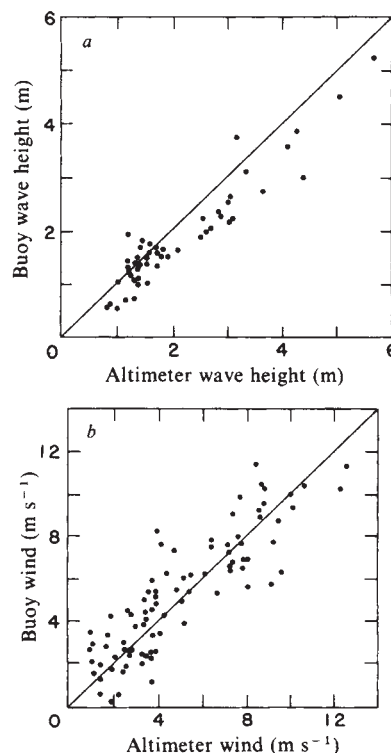


Fig. 2 *a*, Comparison of significant wave height estimated by the Seasat altimeter against 51 buoy measurements. *b*, Comparison of wind speed estimated by the Seasat altimeter against 87 buoy measurements. Data from ref. 3.

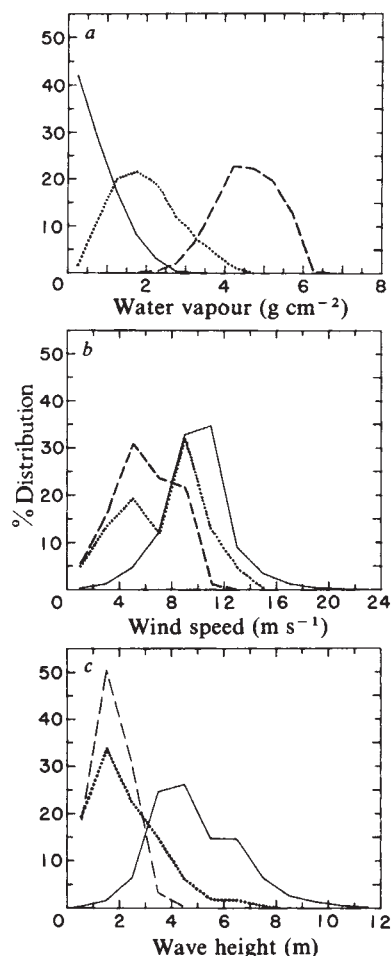


Fig. 3 Distributions of *a*, water vapour; *b*, wind speed; and *c*, wave height estimates from Seasat for the 5° lat. bands centred at 52.5° N (dotted curves), 2.5° N (dashed curves) and 52.5° S (solid curves). All observations between 7 July 1978 and 10 October 1978 are included.

An additional measurement required in any precision altimetric mission is the integrated atmospheric water vapour between the sea surface and the satellite. Water vapour estimates are necessary to make pathlength and attenuation corrections to the ALT radar pulse. On Seasat, integrated water vapour was estimated using the near nadir cell of the SMMR. The presence of water vapour between the satellite and the sea surface increases the atmospheric index of refraction which increases the effective radio frequency path length by 10–50 cm. The altimeter pathlength correction was estimated directly by a linear function of the SMMR brightness temperatures at 18, 21 and 37 GHz (both horizontal and vertical polarization). These pathlength correction estimates were then converted to total integrated atmospheric water vapour based on comparison with direct water vapour measurements from radiosondes. A detailed discussion of this procedure is given in ref. 5.

Cover picture

The 3.5-month Seasat climatological averages of integrated atmospheric water vapour, wind speed and wave height are shown on the cover. More than 3.5 million observations have been averaged globally into 2.5° square areas. The typical number of observations per 2.5° square is around 500 (higher at high latitudes where the satellite ground tracks converge). During the last month of the mission, Seasat was held in a locked orbit with 900-km equatorial spacing of ground tracks over an ~3-day repetition period. This resulted in an uneven sampling distribution over the 2.5° global grid in the 3.5-month averages presented here; 2.5° squares along the locked orbit ground track typically contained nearly twice as many sample data points over

the 3.5-month period as the 2.5° squares between locked orbit ground tracks. (The regions between locked orbit ground tracks were not sampled at all from 9 September to 10 October.) Consequently, to obtain useful pictures of water vapour, wind speed and wave height, the data had to be smoothed zonally with a 7.5° running mean filter (no smoothing meridionally), so the images on the cover effectively represent 3.5-month climatological averages over 2.5° lat. by 7.5° long. areas.

The average water vapour is shown in the upper panel. The highest values are in the tropics (especially the Asian monsoon region and the intertropical convergence zone in the eastern tropical Pacific) and the lowest are at high latitudes. The effects of cool dry air blowing equatorwards over the eastern ocean basins and turning westwards as the trade winds are also evident as well as the region of high water vapour in the southern intertropical convergence zone at about lat. 10° S in the western tropical Pacific. The anomalous observations close to Antarctica are believed to be due to the presence of ice producing increased microwave brightness temperatures.

The average wind speed measured by ALT is shown in the middle panel. Several classical features of the wind field are clear: the north-east and south-east trade winds separated by the doldrums (region of light winds) in both the Atlantic and Pacific Oceans; the horse latitudes (another region of light winds) separating the trade winds from the stormy westerlies in both hemispheres; the south-east trade winds in the south equatorial Indian Ocean and the high wind speeds from the summer monsoon in the northwestern Indian Ocean. The highest wind speeds are in the Southern Hemisphere winter storm region between 50° S and Antarctica (especially south-west of Australia). The narrow region of high wind speeds in the eastern tropical Pacific at about lat. 2° N between long. 90° W and 140° W is a feature previously not seen in ship wind reports which are very sparse in this region. Note also the jet of high wind speed off the coast of California (see ref. 6).

The lower panel shows the average significant wave height during the Seasat mission. This is the first accurate picture of global wave conditions and, we believe, the first reliable measurement by any method of ocean waves for winter season in the Southern Hemisphere. It reveals some interesting features. The average wave height is small in the summer Northern Hemisphere (generally 2–3 m) and the largest waves are located in the winter Southern Ocean where the wind speeds are highest (especially south-west of Australia where mean wave height is >5.5 m). The smallest waves are found in the western Atlantic and Pacific Oceans (average height <1.5 m) in regions of light winds. We expect that further analysis of the ALT wave height data will ultimately lead to a better understanding of the process of wind wave generation.

Discussion

The fact that the 2.1-m contour between 50° and 55° N in the North Pacific does not extend all the way to the coast can probably be explained by the coarse 2.5° lat. by 2.5° long. resolution of the data presented here. Examination of finer-resolution data in this region reveals that the apparent decrease in wave height in the 2.5° averages close to the coast is due to the presence of a distinct wave shadow behind Queen Charlotte Island (the largest of the many islands off the west coast of Canada). In the finer-resolution data, the 2.1-m waves extend all the way to the western shores of these islands but wave heights are <1 m behind the islands.

Note that ice in the ALT footprint causes very erratic behaviour in the shape and power of the returned radar pulse which typically results in rapid changes from very high to very low wave height estimates by ALT. This anomalous behaviour is flagged on the ALT geophysical data records generated from the sensor data by the Jet Propulsion Laboratory, thus providing a potential method for mapping ice coverage from altimeter measurements. Anomalous high wave height estimates have been excluded from our data so that wave height estimates over ice are small. Thus the ice-covered area is represented in the

wave height map on the cover by the light-coloured region close to Antarctica.

While climatological averages emphasize geographical differences within each field, they give no indication of the range of conditions likely to be encountered in a particular region at different times. The rather striking zonal banding of each of the fields suggests that latitudinal distributions can provide some initial insight into water vapour, wind speed and wave height variability during the Seasat mission. The probability distributions of each of the three fields for the three 5° lat. bands centred at 52.5°N, 2.5°N and 52.5°S are shown in Fig. 3.

The features evident in the water vapour distributions in Fig. 3a are not surprising. They clearly show the higher integrated water vapour levels found in the tropics. Only 2% of the observations in the equatorial band were $<3 \text{ g cm}^{-2}$ while 96% of the observations in the high southern latitude band were $<2 \text{ g cm}^{-2}$. This reflects the lower water vapour saturation level of the cold Southern Hemisphere winter air. The cool high northern latitude band shows intermediate water vapour values with 75% of the observations between 1 and 3 g cm^{-2} . A more detailed study of the water vapour estimates might be useful in climatological studies of the global time variability of the latent heat of vapourization transferred from the ocean to the atmosphere.

The wind speed distributions in Fig. 3b show some unexpected features. As the high southern latitude band is in the winter storm region, the higher wind speeds in this band are not surprising. Large variations might also be expected due to the passage of individual storms. However, the winds in this band seem to be relatively steady: 68% of the observations were between 8 and 12 m s^{-1} . In comparison, the trade winds of the equatorial band, which are noted for their steadiness, show a peak at much lower wind speeds but a broader distribution: 76% of the observations were between 4 and 10 m s^{-1} . Another curious feature is the probability distribution for the high northern latitude band: the range of wind speeds is broader than either of the other two bands but there are two pronounced peaks, a primary peak at 9 m s^{-1} and a secondary peak at 5 m s^{-1} . Present studies are aimed at determining whether the double peaks arise from zonal averaging of an inhomogeneous

field or whether they reflect a seasonal change in the wind speed distribution during the 3.5-month period.

The wave height distributions are shown in Fig. 3c. The equatorial band has the narrowest range of wave conditions: 70% of the waves were $<2 \text{ m}$ and virtually none were $>4 \text{ m}$. In comparison, only 50% of the waves in the high northern latitude band were $<2 \text{ m}$ and 5% were $>5 \text{ m}$. In contrast to both the equatorial and high northern latitude bands, the peak in the wave height distribution in the stormy high southern latitude band is at a much higher value (3.5 m as opposed to 1.5 m) and the range of wave conditions is much wider. Only 2% of the waves observed were $<2 \text{ m}$ and over 40% were $>5 \text{ m}$. The secondary peak at 6.5 m is rather puzzling; it is difficult to explain in terms of wind forcing as there is no similar double-peaked structure in the wind speed distribution in this latitudinal band. The double peaks probably result from zonal averaging of the large wave heights south and west of Australia with the smaller wave heights elsewhere in this 5° latitudinal band (see cover).

There is no way of knowing at present whether the 3.5-month average wind speed and wave heights presented here are typical for this time of year. However, these geophysical maps demonstrate the unique ability of satellites to provide global measurements of wind and wave conditions. This will be especially useful in data-sparse regions such as the Southern Hemisphere (see Fig. 1b) where present forecasts of weather and sea states are extremely unreliable. Even in the Northern Hemisphere, forecasting skill is limited in many regions by sparse data coverage over the oceans. An operational altimetric mission could improve worldwide sea state forecasting.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. L. Fedor and G. Brown made their results (Fig. 2) available before publication of their own work.

Received 15 July; accepted 12 October 1981.

1. *Science* **204**, 1405–1424 (1979).
2. Walsh, J., Uliana, E. A. & Yaplee, B. S. *Bound. Layer Met.* **13**, 263–276 (1978).
3. Fedor, L. S. & Brown, G. S. *J. geophys. Res.* (in the press).
4. Brown, G. S. *J. geophys. Res.* **84**(B8), 3974–3978 (1979).
5. Tapley, B. D., Lundberg, J. B. & Born, G. H. *J. geophys. Res.* (in the press).
6. Nelson, C. S. *NOAA Tech. Rep. NMFS SSRF-714* (1977).

Dipoles of the α -helix and β -sheet: their role in protein folding

Wim G. J. Hol*, Louis M. Halie* & Christian Sander†

* Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

† Biophysics Department, Max Planck Institute of Medical Research, Jahnstrasse 29, Heidelberg, FRG

As a result of the regular arrangement of peptide dipoles in secondary structure segments and the low effective dielectric constant in hydrophobic cores, the electrostatic energy of a protein is very sensitive to the relative orientation of the segments. We provide here evidence that the alignment of secondary structure dipoles is significant in determining the three-dimensional structure of globular proteins.

It has long been known that in the α -helix the alignment of the peptide dipoles parallel to the helix axis gives rise to a macrodipole of considerable strength (for review see ref. 1). With a dipole moment for each peptide unit of $\sim 3.5 \text{ D}$, a helix of, for example, 10 residues has a dipole moment of 34 D , as 97% of the peptide dipole points in the direction of the helix axis. It has been shown² for points near the helix termini that the effect of the helix dipole is equivalent to the effect of half a positive unit charge at the N-terminus of the helix and half a negative charge at the C-terminus. The same authors provided evidence that the considerable electric field due to the helix

dipole moment is used by proteins in: (1) binding negatively-charged groups, such as phosphate groups; (2) rendering protein side chains located near the N-terminus more nucleophilic; and, (3) stabilizing charged transition states or intermediates along the catalytic pathway. So far ~ 20 helices have been observed which bind negatively-charged groups near their N-termini. Another 10 proteins have their active site close to the N-terminus of a helix (W.G.J.H., unpublished results). For papain, an extensive quantum mechanical calculation³ showed that the field generated by the helix dipole is a major factor in the stabilization of the sulphhydryl-imidazole ion pair essential in

the catalytic mechanism of this enzyme⁴. Secondary structure formation also seems to be influenced by side chain-main chain electrostatic interactions^{5,6}.

Here we extend the role of the helix dipole to the area of protein folding. In this we are motivated by the idea that the directionally sensitive and relatively long-range dipole-dipole interaction is an important factor in ordering the secondary structure in the hydrophobic interior of folded globular proteins.

Model calculations of helix-helix interaction

Figure 1 shows, for a simple case, the result of model calculations of the electrostatic energy between the backbone dipoles of two helices as a function of their distance and relative orientation. The strongest variation in energy occurs when one helix axis is turned by 180° relative to the other (Fig. 1b). The energy difference between a parallel helix pair and an antiparallel pair is considerable (Fig. 1b,c). These calculations

also show that the interaction between two helix dipoles reaches an upper limit at a helix length of ~35 residues (Fig. 1d). We consider first the extent to which proteins containing helices and no β -sheets reflect these simple electrostatic model considerations. Proteins containing β -structures will be discussed later.

All-helical proteins

In an elegant analysis of protein structures, Levitt and Chotia⁷ showed schematically three major types of all-helical structures. By assigning 'plus' signs to the N-termini and 'minus' signs to the C-termini (indicating the effective charge) in their two-dimensional drawings, the alternating pattern of plus and minus signs tentatively confirms the trend of our model calculations.

Estimates of the actual three-dimensional interaction energies between the helices in the all-helical and predominantly helical proteins tobacco mosaic virus (TMV)^{8,9}, myoglobin¹⁰ and parvalbumin¹¹ are given in Table 1. These calculations show

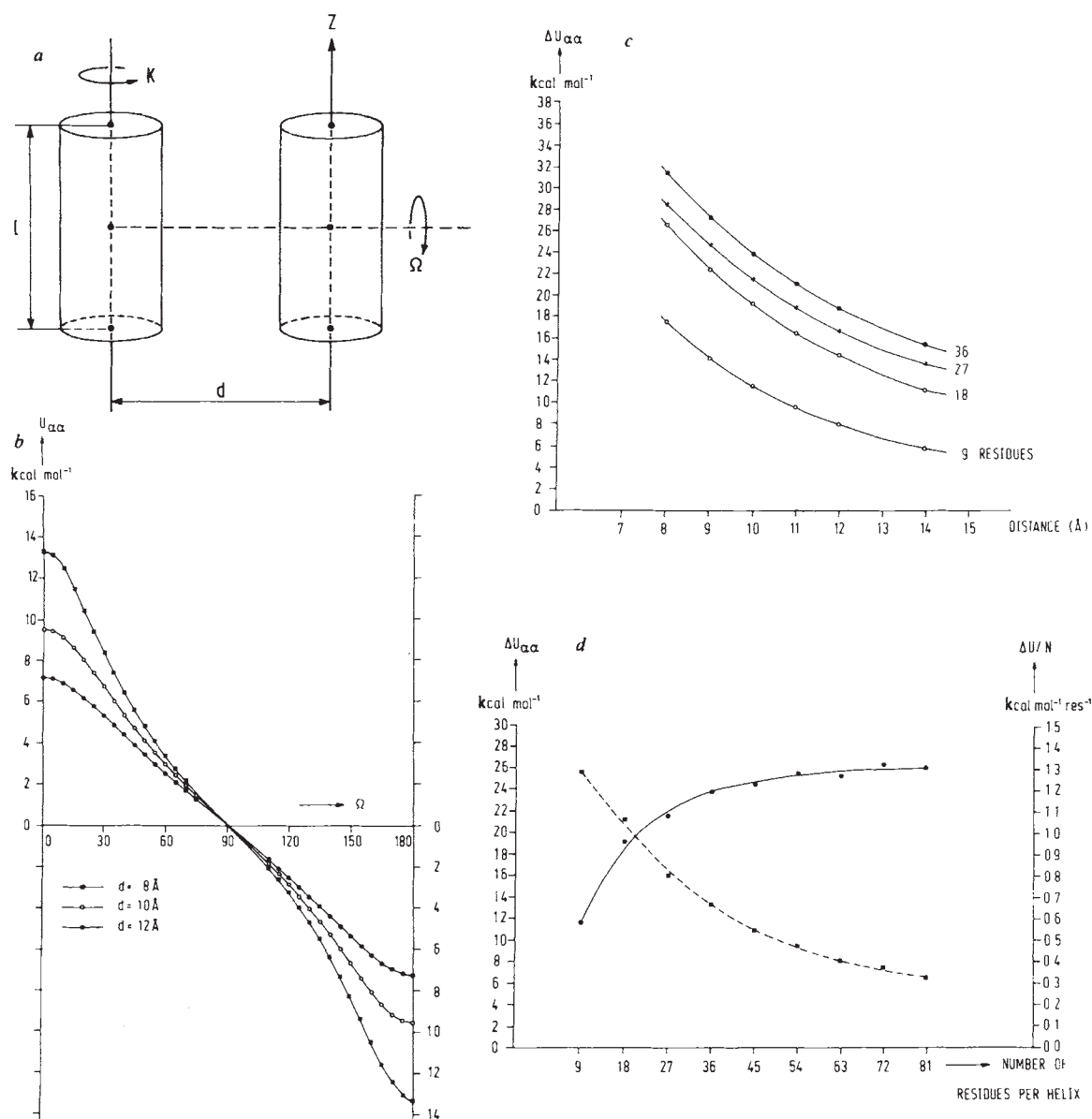


Fig. 1 *a*, The relative position and orientation of two helices in the simple case where the helix axes are perpendicular to the line connecting the centres. A helix can be rotated about its own axis by an angle K (with negligible variation in electrostatic energy) and about the inter-helical axis by an angle Ω (with considerable variation in electrostatic energy). *b*, The strong variation in electrostatic energy between two model α -helices as a function of twist angle Ω . The length of both helices is five turns, that is, 18 residues or 27 Å. *c*, The difference in electrostatic energy between a pair of parallel ($\Omega = 0$) and a pair of antiparallel ($\Omega = 180$) helices falls off gradually with increasing distance, d . *d*, The difference in electrostatic energy between a pair of parallel ($\Omega = 0$) and a pair of antiparallel ($\Omega = 180$) helices at a mutual distance of 10 Å as a function of helix length (solid line). The figure illustrates that the difference becomes less important on a per residue basis for longer helices (broken line). For the model calculations, we used polyaniline in a helical geometry, generated by the program of Hermans *et al.*⁷⁵. The only partial charges are those of the peptide group²: C, 0.42; O, -0.42; N, -0.20 and H, 0.20 unit charge. No cutoff radius was used and dielectric screening was neglected.

Table 1 Calculated electrostatic interaction energies between the backbones of α -helices and β -strands in various proteins

Protein	Type of protein	No. of α -helices	No. of β -strands	$U_{\alpha\alpha}$ (kcal mol ⁻¹)	$U_{\alpha\beta}$ (kcal mol ⁻¹)	Ref.
Met-myoglobin	α_a	9	—	-22.7	—	10
TMV	α_a	4	—	-14.3	—	8, 9
Parvalbumin	α_a	6	—	-11.8	—	11
Cytochrome <i>b₅</i>	α_a	6	—	-4.7	—	22
Phospholipase A2	$\alpha_a\beta_a$	4	2	-8.3	-0.8	41
Cytochrome <i>c</i>	$\alpha_a\beta_a$	4	3	-3.4	-1.4	42
Hen egg-white lysozyme	$\alpha_a\beta_a$	6	5	-12.8	-1.6	19
Phage T4 lysozyme	$\alpha_a\beta_a$	10	3	-23.5	-0.7	20
Thermolysin	$\alpha_a\beta$	7	14	-3.6	-5.4	21
Lactate dehydrogenase	$\alpha_p\beta_p$	9	15	+14.0	-14.0	43
Alcohol dehydrogenase	$\alpha_p\beta_p$	10	20	+14.2	-15.3	44
Glyceraldehyde phosphate dehydrogenase	$\alpha_p\beta_p$	7	18	+5.8	-5.7	45
Adenylate kinase	$\alpha_p\beta_p$	10	5	+4.9	-10.5	46
Rhodanese	$\alpha_p\beta_p$	10	10	+13.5	-10.3	32
Subtilisin	$\alpha_p\beta_p$	8	9	+5.2	-15.0	47
Dihydrofolate reductase	$\alpha_p\beta_p$	4	8	+3.2	-3.8	48
Flavodoxin	$\alpha_p\beta_p$	5	7	+13.3	-4.9	36
Triose phosphate isomerase	$\alpha_p\beta_p$	8	8	+6.6	-7.0	49
Carboxypeptidase	$\alpha_p\beta_p$	9	12	+11.4	-11.0	50
Papain/actinidin	$\alpha\beta+\beta_a$	6	8	+13.3	-6.6	51, 52

The table demonstrates the favourable alignment of α -helices in all- α proteins and the compensation for unfavourable helix interactions by favourable helix-strand interactions in other proteins. $U_{\alpha\alpha}$ is the sum of all main-chain dipole helix-helix interactions, $U_{\alpha\beta}$ the sum of all main-chain dipole helix-strand interactions. $U = \sum (q_i q_j / r_{ij})$, where r_{ij} is the distance between two charges q_i and q_j , with no cut-off in r and with the partial charges: O = -0.42, C = +0.42, N = -0.20, H = +0.20, corresponding to a peptide dipole moment of 3.5 D in the direction of the N-H and O-C bonds. Indices i and j label the residues in the segments; i and j are not in the same segment. A dielectric constant was not used because of the breakdown on the atomic scale of continuum theories on which the concept of dielectric constant is based. This breakdown is illustrated quantitatively by molecular dynamics simulations⁵³. The dielectric response, however, can be estimated and is physically due to the electronic polarizability of the protein atoms and the orientational polarizability of solvent water. The effects of the former can be separated into two contributions which occur in opposite directions. (1) Recent theoretical estimates of an effective dielectric constant for the non-polar atoms in the protein interior are $\epsilon = 2$ (for example, derived from atomic polarizabilities⁵⁴). This decreases U by a factor of ~ 2 . (2) Cooperatively aligned backbone dipoles in both α -helix and β -sheet mutually polarize each other, increasing the O $\leftarrow$ C and N $\rightarrow$ H dipoles by a factor of ~ 1.4 (from 3.5 to 4.8–5.0 D; see refs 1, 55) and thus increasing the interaction energies U by a factor of $1.4 \times 1.4 = 2$. The dielectric screening due to water is difficult to calculate. Kirkwood and Westheimer⁵⁶ and Ehrenson⁵⁴ have estimated the effective screening of charge-charge and charge-dipole interactions in the interior of a sphere of low dielectric surrounded by water: the screening is only effective near the surface and dies off quickly in the interior. As helices are usually nearer the protein surface than parallel β -sheets, this means that our $U_{\alpha\alpha}$ values are an overestimate relative to the $U_{\alpha\beta}$ values. Overall, these estimates of the dielectric response indicate that our energy estimates of $U_{\alpha\alpha}$ and $U_{\alpha\beta}$ are incorrect by no more than a factor of 2. The coordinates are from the Protein Data bank⁵⁷, except for dihydrofolate reductase, TMV and phospholipase, the coordinates of which were provided by J. Kraut, A. Bloomer and J. Drenth, respectively. The position of the peptide hydrogen was generated according to standard geometry. The lengths of the helices and strands were taken from Levitt and Greer⁵⁸, except in the following cases: TMV⁸, phospholipase⁴¹, lactate dehydrogenase⁴³ and actinidin⁵² were taken from the literature and dihydrofolate reductase⁴⁸ was obtained by inspection on an interactive display. Exceptionally favourable interactions between secondary structure segments next to each other in the sequence were removed by deleting residues from the calculations so that no atoms from different segments were closer than 3.6 Å. Parts of kinked helices were merged into one helix. The types of proteins are grouped according to the predominant arrangement of helices and strands. The plus sign in the second column indicates that different domains belong to different types of arrangement. For papain⁵¹ and the closely related protease actinidin⁵² there is no obvious cancellation of the parallel helix dipoles by either strands or by many exposed charged residues. There is, however, an unusual placement of charged interior residues and a network of interior water molecules⁵². Therefore, the helical domain of these proteins is intriguing from the electrostatic as well as from the hydrophobic viewpoint.

very favourable overall electrostatic interactions between the helices, in complete qualitative agreement with our model calculations. Although these energies are difficult to calculate precisely (see Table 1 legend), we do not expect any change in the qualitative conclusions when more accurate calculations are available.

The all-helical proteins myohaemerythrin¹², ferritin¹³, cytochrome *b₅₆₂* (ref. 14) and cytochrome *c'* (ref. 15) strongly resemble the optimally antiparallel TMV structure¹⁶. The all-helical protein uteroglobin contains only antiparallel helices¹⁷. The larger protein cytochrome *c* peroxidase¹⁸ contains eight helices of which five form an antiparallel arrangement and none are aligned in parallel. Our model is also supported by the structure of 6-phosphogluconate dehydrogenase (M. Adams and S. White, in preparation): four pairs of antiparallel helices are wrapped perpendicularly around two long antiparallel helices which form the core of this protein. In addition, helices in the helical domains of hen egg-white lysozyme¹⁹, phage T4 lysozyme²⁰, thermolysin²¹ and cytochrome *b₅* (ref. 22) have a definite tendency to run antiparallel. This is reflected in the negative interaction energies $U_{\alpha\alpha}$ calculated for these proteins (see Table 1).

Proteins with alternating α and β structures

From an electrostatic viewpoint, it is surprising that in proteins with alternating α -helices and β -strands, the helices are roughly parallel, that is, with seemingly unfavourable helix dipole alignment. In studying this phenomenon electrostatically, we examined the arrangement of peptide units in parallel β -strands. From schematic drawings of hydrogen bonding patterns in parallel β -sheets (Fig. 2), it is obvious that the N $\rightarrow$ H as well as the C $\leftarrow$ O dipoles point backwards with respect to the N- to C-terminal direction of the strands. This suggests that the

parallel β -sheet has a significant overall dipole moment, with the N-terminal end of the strands corresponding to the positive end of the dipole. Quantitatively, about one-third of the peptide dipole is parallel to the strand direction. This allows an approximate description of a parallel β -strand as having 1/15 of a positive charge near the C-terminus (Fig. 2a). Electrostatically, then, the α - β dipolar interaction is favourable when helices and strands are antiparallel.

A second aspect of the electrostatic interaction between helices and sheets concerns the twist of the sheet²³. A consequence of this twist is that the 'parallel' helices in the α/β proteins make considerable angles (typically -40° ; ref. 24) with each other, thus decreasing appreciably the unfavourable interaction between their dipoles (Fig. 1b).

We have calculated the actual α - α and α - β dipole interactions in various α/β proteins and the simple considerations above seem to be valid generally (Table 1). The negative values for $U_{\alpha\beta}$ qualitatively confirm our hypothesis of a favourable strand-helix interaction. This interaction compensates fully or partly for the unfavourable helix-helix interaction (Table 1). In our energy estimates, we disregarded the difference in electrostatic screening near the surface and in the interior. As the parallel β -sheets are sandwiched by the helices, we have overestimated $U_{\alpha\alpha}$ relative to $U_{\alpha\beta}$ and so the actual total energies are probably more favourable. Thus, domain structures with parallel central β -sheets antiparallel to surrounding helices may be significantly stabilized by electrostatic interactions.

All- β proteins

In globular proteins containing only β -sheets, there is an overwhelming tendency for the strands to align in an antiparallel manner (ref. 7 and Table 2). The same seems to be true for fibrous β -sheet proteins^{25,26}. Electrostatically, an energy

difference between the two kinds of pleated sheet may be related to the fact that in the antiparallel sheet all peptide dipole moments seem to cancel each other out: there is no residual moment in either direction (Fig. 2b). In the parallel arrangement the components of the dipole moment parallel to the strand direction interact unfavourably with each other. A simple model calculation, assuming only a dipole moment in the parallel strand and no dipole moment in the antiparallel strand, shows that this effect may be $\sim 0.4 \text{ kcal mol}^{-1}$ for a pair of short β -strands containing four residues. This gives a difference in electrostatic energy of $\sim 0.8 \text{ kcal mol}^{-1}$ between a parallel and antiparallel arrangement of three strands, which is much smaller than the energy difference for α -helices, but it may nevertheless be significant.

The $\alpha\beta$ and larger proteins

A small number of proteins do not fall into the all- α , α/β or all- β categories, for example the C-terminal fragment of the ribosomal protein L7/L12 (ref. 27), in which the three helices and three β -strands are optimally antiparallel to each other. The L7/L12 protein is very stable²⁷—probably due not only to the close packing of nonpolar residues in its unusually hydrophobic core, but also to the optimally antiparallel alignment of secondary structure backbones embedded within that core. Proteins like L7/L12 can be described as $\alpha\beta_a$, whereas the 'classical' α/β proteins may be coined $\alpha_p\beta_p$ (a, antiparallel; p parallel). The all-helical proteins may then be described as α_a , and the all- β proteins as β_a .

In larger proteins a mixture of folding patterns occurs, for example, in *p*-hydroxybenzoate hydroxylase²⁸. This protein consists of three domains which can be described as $\alpha_p\beta_p$, β_a and α_a .

In non-compact interfaces between subunits in oligomeric proteins, dielectric screening by the solvent will be much greater than in hydrophobic cores. Hence, preferred orientations for α - and β -dipoles from different subunits are expected to be less pronounced.

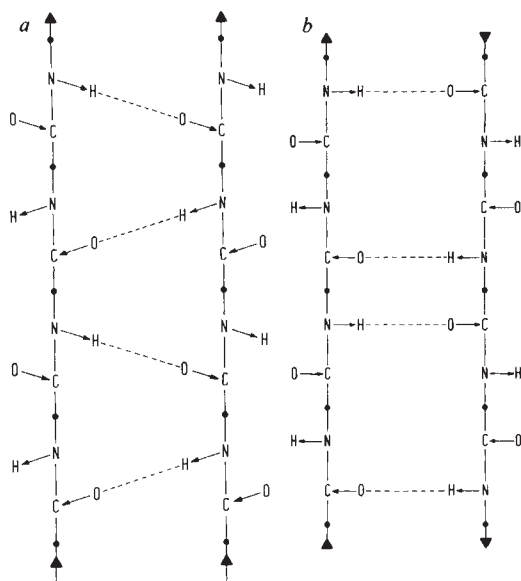


Fig. 2 Schematic representation of the arrangement of the N-H and C=O dipoles in β -sheets. *a*, Pair of parallel β -strands; *b*, pair of antiparallel β -strands. In parallel and antiparallel strands the N-H and O=C bonds have different directions relative to the strand direction. In the antiparallel case the bonds are fairly perpendicular to the strand direction, with little or no resultant dipole moment. In the parallel case, these bonds make angles of $\sim 20^\circ$ with the N to C strand direction, thus about one-third of the peptide dipole moment runs parallel to the strand. Hence the component along the strand direction is $\mu = 1/3 \mu(\text{peptide}) = 1/3 \times 3.5 \text{ D} = 1.15 \text{ D} = 0.23 \text{ eÅ}$. As the repeat distance along a β -strand is $\sim 3.5 \text{ Å}$ [refs 25, 76], a parallel β -strand can be considered as a dipole with a charge of $+1/15$ at its N-terminus and $-1/15$ at its C-terminus. Similar reasoning shows that the dipole of the α -helix can be approximated by a charge couple of $+1/2$ and $-1/2$, respectively, at the N and C ends². Thus on a per length basis, the β -parallel strand dipole is about a factor of five weaker than the helix dipole.

Table 2 The predominance of antiparallel over parallel strand pairs in all- β proteins and all- β domains

Protein	Strand pairs		Ref.
	β_a	β_p	
Elastase	11	0	59
Papain (domain II)	6	0	51
Ribonuclease S	3	0	60
Concanavalin A	12	0	61, 62
Rubredoxin	2	0	63
High-potential iron protein	2	0	64
IgG (Fab')	14	1	65
Prealbumin	5	1	66
Neurotoxin	4	0	67, 68
Superoxide dismutase	6	0	69
Bacteriochlorophyll protein	16	1	70
Subtilisin inhibitor	4	0	71
Soybean trypsin inhibitor	3	0	72
Tomato bushy stunt virus	14	0	73
Acid proteases	7	2	74
Total	109	5	

All- β proteins here includes those with a negligible number of helices.

Charged side chains

In addition to the interactions between secondary structure dipoles in the protein interior, electrostatic interactions of charged side chains with each other and with the α -helix dipoles are another component of the total electrostatic energy of a protein molecule. This component will, in most cases, be less important than the interactions between the α and β dipoles. Experimental evidence for this is provided by chemical modification studies in which the charge of exposed residues is removed or even reversed while the catalytic activity of the proteins is little affected²⁹⁻³¹. In addition, comparison of different amino acid sequences which form strikingly similar structures shows little similarity in the position of residues of similar charge (for example, rhodanese³² and the serine proteases³³). A likely explanation for this is effective electrostatic screening of charged residues due to counterions and solvent dipoles^{34,35}.

Flavodoxin probably has a significant electrostatic interaction between charged residues and α -helix dipoles³⁶. In this protein, the charged residues are distributed non-uniformly with respect to the helices: ~ 10 negative charges occur near the N-termini of the helices without any oppositely charged side-chain partner nearby. The helices in this α/β protein interact quite unfavourably (Table 1), and in this case the large number of favourable charge-helix dipole interactions compensates, together with the favourable α - β dipole interactions, for the unfavourable helix dipole interactions.

The tropomyosin dimer, which is not a globular protein but forms a coil of two very long parallel helices of 284 residues³⁷, shows the predominance of a large number of salt bridges over the interaction between the helix dipoles. As McLachlan and Stewart observed³⁷, the charged residues in this protein are situated in such a way that like charges clash when forming antiparallel helices, but as many as 30 salt bridges can be formed in the parallel arrangement. As the helix dipole interaction does not increase after ~ 30 residues (Fig. 1d), the charge-charge interaction is now the major factor in determining the direction of the helices with respect to each other. However, α -helical homopolypeptides, which lack such charge pairs (for example, polyalanine) actually seem to prefer an antiparallel arrangement, both in fibres and in lamellar single crystals^{26,38}, which is in complete agreement with the results of our model calculations (Fig. 1).

Conclusion

If the values of the electrostatic energies given in Table 1 are compared with the experimentally observed values of $12 \pm 5 \text{ kcal mol}^{-1}$ for the free energy difference between folded and

unfolded protein conformations³⁹, it seems that the interaction between the main chain dipole moments of secondary structure elements has a significant role in determining the conformation of a protein molecule.

Our electrostatic considerations are also in accord with the frequent occurrence of $\alpha\alpha$, $\beta\beta$ and $\beta\alpha\beta$ units in proteins⁷. The picture which emerges is that the interior of globular proteins is a medium with low dielectric screening in which the interactions between secondary structure dipoles lead to preferred folding patterns. Kauzmann⁴⁰ suggested that hydrogen bonds between peptide links and hydrophobic interactions are the most important factors in determining the overall configuration of

globular protein molecules. It seems reasonable to suggest that relatively long-range electrostatic interactions in the hydrophobic protein interior should be added to this list.

We thank P. van Duijn and J. Drenth for stimulating and critical discussions, E. Samulski for his contributions in the early stages of this work, the Groningen University Computer Centre for their cooperation, and C. Moser at CECAM, Orsay, for the organization of workshops, which brought together minds with similar interests. This research was carried out, in part, under the auspices of the Dutch Organisation of Chemical Research (SON), and supported in part by a grant from the Deutsche Forschungsgemeinschaft.

Received 15 May; accepted 5 October 1981.

1. Wada, A. *Adv. Biophys.* **9**, 1-63 (1976).
2. Hol, W. G. J., Van Duijn, P. Th. & Berendsen, H. J. C. *Nature* **273**, 443-446 (1978).
3. Van Duijn, P. Th., Thole, B. T. & Hol, W. G. J. *Biophys. Chem.* **9**, 273-280 (1979).
4. Drenth, J., Swen, H. M., Hoogenstraaten, W. & Sluiterman, L. A. *Æ. Proc. K. ned. Akad. Wet.* **C78**, 104-110 (1975).
5. Jernigan, R. L., Miyazawa, S. & Szu, S. C. *Macromolecules* **13**, 518-525 (1980).
6. Blagdon, D. E. & Goodman, M. *Biopolymers* **14**, 241-245 (1975).
7. Levitt, M. & Chothia, C. *Nature* **261**, 552-558 (1976).
8. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, R. *Nature* **276**, 362-368 (1978).
9. Stubbs, G., Warren, S. & Holmes, K. *Nature* **267**, 216-221 (1977).
10. Kendrew, J. C. *et al.* *Nature* **185**, 422-427 (1960).
11. Kretsinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248**, 3313-3326 (1973).
12. Hendrickson, W. A. & Ward, K. B. *J. biol. Chem.* **252**, 3012-3018 (1977).
13. Banyard, S. H., Stammers, D. K. & Harrison, P. M. *Nature* **271**, 282-284 (1978).
14. Matthews, F. S., Bethge, P. H. & Czerwinski, E. W. *J. biol. Chem.* **254**, 1699-1706 (1979).
15. Weber, P. C. *et al.* *Nature* **286**, 302-304 (1980).
16. Weber, P. C. & Salemm, R. *Nature* **287**, 82-84 (1980).
17. Mornon, J. P., Fridlansky, F., Bally, R. & Milgrom, E. *J. molec. Biol.* **137**, 415-429 (1980).
18. Poulos, Th. L. *et al.* *J. biol. Chem.* **255**, 575-580 (1980).
19. Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. in *The Enzymes* Vol. 7 (ed. Boyer, P. D.) 666-686 (Academic, New York, 1972).
20. Remington, S. J. *et al.* *J. molec. Biol.* **118**, 81-98 (1978).
21. Colman, P. M., Jansonius, J. N. & Matthews, B. W. *J. molec. Biol.* **70**, 701-724 (1972).
22. Matthews, F. S., Levine, M. & Argos, P. *J. molec. Biol.* **64**, 449-464 (1972).
23. Chothia, C. *J. molec. Biol.* **75**, 295-302 (1973).
24. Janin, J. & Chothia, C. *J. molec. Biol.* **143**, 95-128 (1980).
25. Arnott, S., Dover, S. D. & Elliot, A. *J. molec. Biol.* **30**, 201-208 (1967).
26. Fraser, R. D. B. & MacRae, T. P. *Conformation in Fibrous Proteins* (Academic, New York, 1973).
27. Leijonmarck, M., Eriksson, S. & Liljas, A. *Nature* **286**, 824-826 (1980).
28. Wierenga, R. K., De Jong, R. J., Kalk, K. H., Hol, W. G. J. & Drenth, J. *J. molec. Biol.* **131**, 55-73 (1979).
29. Cohen, L. A. in *The Enzymes* Vol. 1 (ed. Boyer, P. D.) 147-211 (Academic, New York, 1970).
30. Ottesen, M., Johansen, J. T. & Swendsen, I. in *Structure-Function Relationships of Proteolytic Enzymes* (eds Desnuelle, P., Neurath, H. & Ottesen, M.) 175-187 (Munksgaard, Copenhagen, 1970).
31. Glazer, A. N. in *The Proteins* Vol. 2 (eds Neurath, H. & Hill, R. L.) 1-103 (1976).
32. Ploegman, J. H., Drent, G., Kalk, K. H. & Hol, W. G. J. *J. molec. Biol.* **123**, 557-594 (1978).
33. Delbaere, L. T. J., Hatchoon, W. L. B., James, M. N. G. & Thiessen, W. E. *Nature* **257**, 758-763 (1975).
34. Ehrenson, S. *J. Am. chem. Soc.* **98**, 7510-7514 (1976).
35. Rees, D. C. *J. molec. Biol.* **141**, 323-326 (1980).
36. Burnett, R. M. *et al.* *J. biol. Chem.* **249**, 4383-4392 (1974).
37. McLachlan, A. D. & Stewart, M. *J. molec. Biol.* **103**, 271-298 (1976).
38. Arnott, S. & Wonacott, A. *J. molec. Biol.* **21**, 371-383 (1966).
39. Privalov, P. L. *Adv. Protein Chem.* **33**, 167-241 (1979).
40. Kauzmann, W. *Adv. Protein Chem.* **14**, 1-63 (1959).
41. Dijkstra, B. W., Drenth, J., Kalk, K. H. & Vandermaelen, P. J. *J. molec. Biol.* **124**, 53-60 (1978).
42. Takano, T. *et al.* *J. biol. Chem.* **252**, 776-785 (1977).
43. Holbrook, J. J., Liljas, A., Steindel, S. J. & Rossmann, M. G. in *The Enzymes* Vol. 11 (ed. Boyer, P. D.) 191-292 (Academic, New York, 1975).
44. Brändén, C.-I., Jörnvall, H., Eklund, H. & Furugren, B. in *The Enzymes* Vol. 11 (ed. Boyer, P. D.) 103-190 (Academic, New York, 1975).
45. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. *J. molec. Biol.* **82**, 563-585 (1974).
46. Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. *Nature* **250**, 120-123 (1974).
47. Wright, C. S., Alden, R. A. & Kraut, J. *Nature* **221**, 235-242 (1969).
48. Matthews, D. *et al.* *J. biol. Chem.* **253**, 6946-6954 (1978).
49. Banner, D. W. *et al.* *Nature* **255**, 609-614 (1975).
50. Hartsuck, J. A. & Lipscomb, W. N. in *The Enzymes* Vol. 3 (ed. Boyer, P. D.) 1-56 (Academic, New York, 1971).
51. Drenth, J., Jansonius, J. N., Koekoek, R. & Wolthers, B. G. in *The Enzymes* Vol. 3 (ed. Boyer, P. D.) 485-499 (Academic, New York, 1971).
52. Baker, E. N. *J. molec. Biol.* **141**, 441-484 (1980).
53. Pollock, E. L., Alder, B. J. & Pratt, L. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 49-51 (1980).
54. Pethig, T. *Dielectric and Electronic Properties of Biological Materials* (Wiley, New York, 1979).
55. Applequist, J. & Mahr, T. G. *J. Am. chem. Soc.* **88**, 5419 (1966).
56. Kirkwood, J. G. & Westheimer, F. H. *J. chem. Phys.* **6**, 5419-5429 (1938); **7**, 437 (1939).
57. Bernstein, F. C. *et al.* *J. molec. Biol.* **112**, 535-542 (1977).
58. Levitt, M. & Greer, J. *J. molec. Biol.* **114**, 181-239 (1977).
59. Sawyer, L. *et al.* *J. molec. Biol.* **118**, 137-208 (1978).
60. Wyckoff, H. W. *et al.* *J. biol. Chem.* **245**, 305-328 (1970).
61. Hardman, K. D. & Ainsworth, C. F. *Biochemistry* **11**, 4910-4919 (1972).
62. Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* **250**, 1525-1547 (1975).
63. Watenpugh, K. D., Sieker, L. C., Herriott, J. C. & Jensen, L. H. *Acta crystallogr.* **B29**, 943-956 (1973).
64. Carter, C. W. *et al.* *J. biol. Chem.* **249**, 4212-4225 (1974).
65. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3440-3444 (1974).
66. Blake, C. C. F., Geisow, M., Oatley, S. J., Rérat, B. & Rérat, C. *J. molec. Biol.* **121**, 339-356 (1978).
67. Tsernoglou, D. & Petsko, G. A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 971-974 (1977).
68. Low, B. W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **73**, 2991-2994 (1976).
69. Richardson, J. S., Thomas, K. A., Rubin, B. H. & Richardson, D. C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1349-1353 (1975).
70. Matthews, B. W., Fenna, R. A., Bolognesi, M. C., Schmid, M. F. & Olson, J. M. *J. molec. Biol.* **131**, 259-285 (1979).
71. Mitsui, Y., Satow, Y., Watanabe, Y. & Iitaka, Y. *J. molec. Biol.* **131**, 697-724 (1979).
72. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. *Biochemistry* **13**, 4212-4228 (1974).
73. Harrison, S. C., Olson, A. J., Schutt, C. E. & Winkler, F. K. *Nature* **276**, 368-378 (1978).
74. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A. & Blundell, T. L. *Nature* **271**, 618-621 (1978).
75. Hermans, J. & McQueen, J. E. *Acta crystallogr.* **A30**, 730-739 (1974).
76. Pauling, L. & Corey, R. B. *Proc. natn. Acad. Sci. U.S.A.* **39**, 253-256 (1953).

Clustered arrangement of immunoglobulin λ constant region genes in man

Philip A. Hieter*, Gregory F. Hollis*, Stanley J. Korsmeyer†, Thomas A. Waldmann† & Philip Leder*

* Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, and † Metabolism Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

The immunoglobulin λ light chain locus of man contains six λ -like genes arranged tandemly on a 50-kilobase segment of chromosomal DNA. The sequences of three of these genes correspond to three known non-allelic λ chain isotypes: Mcg, Ke⁻Oz⁻ and Ke⁻Oz⁺. They surround a highly polymorphic and evidently unstable region that is repeatedly deleted when cloned in Escherichia coli. Three additional, but as yet unlinked, λ -like sequences have also been cloned, suggesting that the λ genes form an unexpectedly large family within the human genome.

THE formation of an active immunoglobulin light chain gene (κ or λ) involves the covalent joining of two distantly encoded segments of germ-line DNA¹⁻⁵. In the case of the κ variable

region of the mouse, V region diversity is accounted for largely by the availability of several hundred germ-line variable genes^{6,7} that can be joined to one of four active J region segments^{8,9}. The

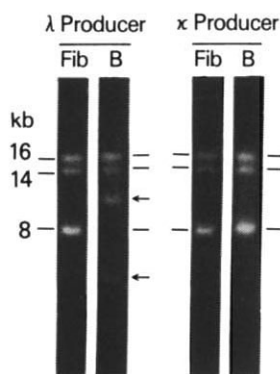


Fig. 1 Analysis of λ gene-bearing *Eco*RI fragments in human B-lymphocyte DNA. Light chain-expressing B cells (B) examined included monoclonal B lymphocytes, expressing λ or κ chain (as indicated) on their cell surface, taken from patients with high-count chronic lymphocytic leukaemia. Fibroblast (Fib) cell cultures served as germ-line controls for each individual. Genomic DNA was extracted from cells²⁵, digested with *Eco*RI restriction endonuclease, size fractionated by agarose gel electrophoresis, transferred to nitrocellulose²⁹, hybridized with ³²P-labelled³⁰ 2.5-kb *Eco*RI–*Hind*III fragment obtained from the Hu λ 5 library clone (see Fig. 2 legend), and autoradiograms developed. DNA fragments identified in fibroblast DNA are indicated by thin lines; rearranged fragments in B-cell DNA by arrows. Identical results were obtained using as probe the subcloned 3.4-kb *Eco*RI–*Hind*III fragment obtained from the Hu λ 5 library clone (see Fig. 2).

precise cross-over point of this V–J joining can itself vary so as to create additional diversity around the site of recombination^{8–10}. The arrangement of the human κ region gene is similar, but includes an additional active J region segment¹¹.

The arrangement of mouse λ genes differs from that of κ genes in that the mouse V_{λ} repertoire is very small¹ and only a single J region segment is likely to be encoded adjacent to each of the four known λ constant region genes^{12,13}. The human λ gene locus probably differs from that of the mouse as human antibody molecules are associated with a diverse population of λ chains, whereas mouse antibodies are not. Such a difference may reflect a much larger human V_{λ} gene repertoire. In addition, we have observed that there is a specific order of light chain gene recombination in which κ precedes λ ^{14,15}. The mechanism of this ordered rearrangement is unknown.

The human locus seems to be particularly favourable for examining the molecular basis for λ gene behaviour. The human λ constant region consists of at least four non-allelic forms that

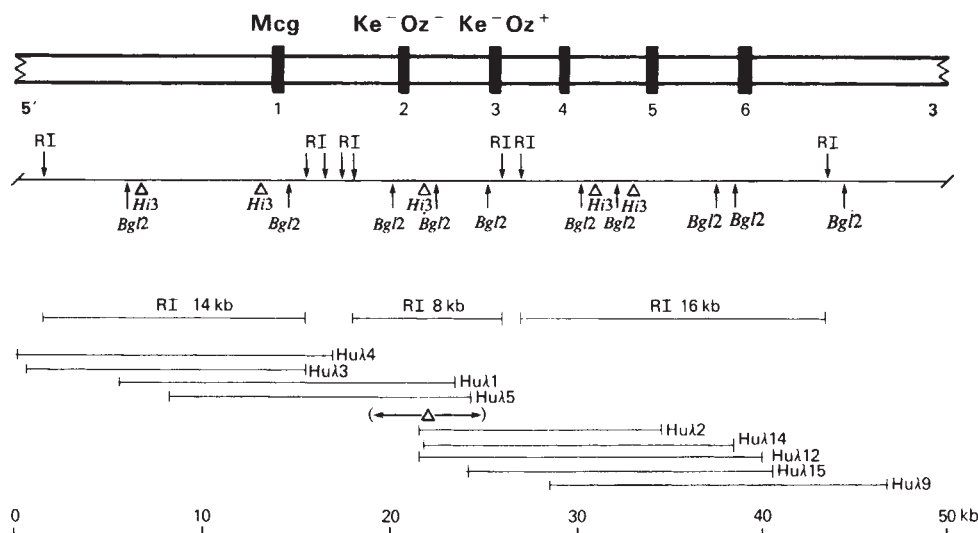
differ by limited amino acid substitutions to produce the serological markers Kern, Oz and Mcg (refs 16–19). Several additional λ variants have been recognized, but it is unknown whether these represent allelic variants or distinct isotypes^{20–22}. Here we report the cloning and initial characterization of the germ-line λ constant region genes of man. These genes constitute an unexpectedly complex locus of at least six non-allelic genes that are arranged tandemly on a single chromosome. In addition, we have identified a polymorphic region within the locus that suggests a segment of particular genetic instability.

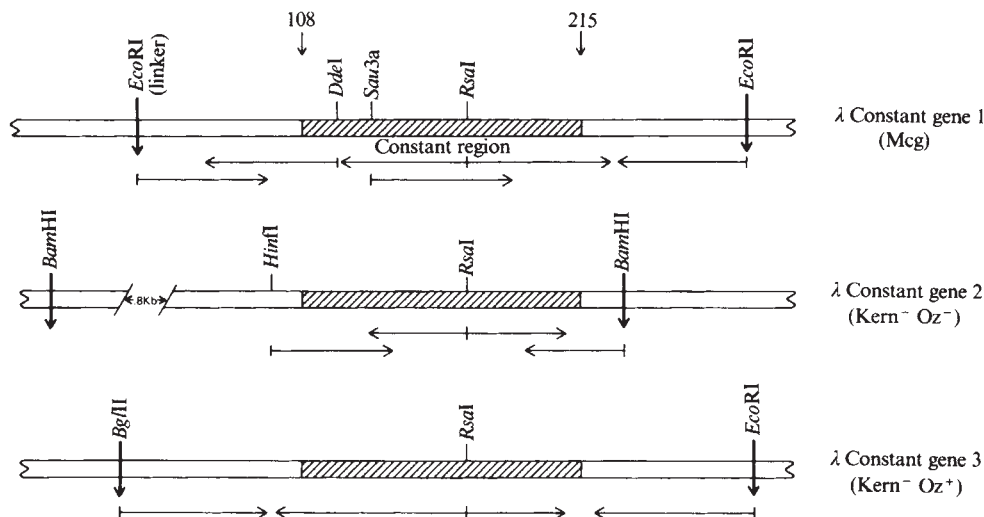
Molecular cloning and identification of the human λ constant region genes

To identify the human λ constant region genes, we used a mouse–human cross-hybridization strategy similar to that described for the identification of the human κ genes¹⁰. Our probe was a 0.24-kilobase (kb) *Mbo*II DNA restriction fragment derived from a mouse λ 1 cDNA clone (p– λ –1; provided by A. Bothwell). This fragment corresponds to the gene segment that encodes amino acids 123 to 203 of the λ 1 constant region. The mouse λ constant region probe was used in conditions of low stringency to screen a human fetal liver library consisting of randomly cleaved DNA fragments cloned in bacteriophage λ CH4A (provided by T. Maniatis)²³. In this way, seven distinct clones were obtained.

If the cloned fragments actually contained λ constant regions, we would expect them to hybridize to rearranged fragments of DNA in λ -producing B cells. We therefore examined DNA derived from λ light chain-bearing B lymphocytes obtained from a patient with high-count chronic lymphocytic leukaemia. (The leukaemic expansion of a lymphocyte is effectively monoclonal.) Subcloned fragments derived from one of the library clones were used as probes (Fig. 1). To control for possible polymorphic differences in the arrangement of hybridizing DNA fragments²⁴, DNA extracted from fibroblast cultures established from the same individual served as a germ-line-like control. As shown in Fig. 1, the cloned probe detected two rearranged fragments in the DNA from the λ -producing B cells, whereas the DNA extracted from control κ -producing B cells demonstrated no such rearrangement. Note that the monoclonal λ -producing cells show rearrangement of two DNA fragments and that the 8-kb germ-line fragment has been lost in these cells. The most likely interpretation of this result is that rearrangement has occurred on both homologous chromosomes and that these rearrangements have occurred within or 3' to the 8-kb *Eco*RI fragment (see map, Fig. 2).

Fig. 2 Physical map of the human λ light chain constant region genes. The arrangement of a cluster of six λ constant genes (filled boxes) along chromosomal DNA is shown at the top. The orientation is 5' to 3', left to right. A detailed map for the restriction enzymes *Eco*RI, *Hind*III and *Bgl*II is shown. The map was deduced by overlapping a series of library clones isolated from a human fetal liver DNA library²³ and comparing cloned, genomic *Eco*RI fragments from that locus. Single and double restriction enzyme digestion patterns for *Eco*RI, *Hind*III, *Bgl*II and *Cfo*I were compared and analysed for the presence of λ gene sequences by Southern blotting. Nine representative library clones (out of a total of 13) are shown. The library clones fell into two groups represented by clones 4, 3, 1, 5 and clones 2, 14, 12, 15, 9. Library clones that spanned the region indicated by Δ invariably deleted extensive regions during propagation in the bacterial host. The unambiguous linkage between the two groups of library clones was obtained by demonstrating that the restriction fragments generated from the cloned genomic 8-kb *Eco*RI fragment were identical to the sum of restriction fragments generated from an *Eco*RI–*Hind*III subclone of Hu λ 5 plus a *Hind*III–*Eco*RI subclone of Hu λ 2. The library clones are represented by thin lines and designated as Hu λ . The genomic *Eco*RI fragments (RI) are represented by thin lines and their fragment lengths given. The three most 5' genes correspond to the non-allelic forms Mcg, Kern–Oz[–] and Kern–Oz⁺, as determined by nucleotide sequence analysis (see Fig. 4).





Six λ -like constant region genes are arranged along a 50-kb segment of chromosomal DNA

Using the same human DNA fragments as probes, we re-screened the human clone library and obtained 19 distinct library clones (7 of which were identical to those obtained with the mouse probe, described above). In addition, the 8-, 14- and 16-kb genomic *EcoRI* fragments identified in Southern blots of genomic DNA (Fig. 1) were cloned directly using a fragment purification approach described previously²⁵. By ordering the overlapping clones from restriction endonuclease site mapping data and Southern blot analysis (see Fig. 1 legend), we constructed the linkage map of the λ constant genes shown in Fig. 2. The map extends over 50,000 base pairs (bp) and contains at least six λ gene-like sequences tandemly arranged along the chromosome. The restriction enzymes *EcoRI*, *HindIII*, *BglII* and *CfoI* were used to establish the map. Thirteen of nineteen library clones overlap this locus; the remaining six clones are as yet uncharacterized. Six regions within this locus demonstrate strong homology by hybridization to the mouse λ constant region probe (data not shown). Note that these regions are regularly spaced along the chromosome ~5,000 bp apart.

Identification of three additional λ -like sequences

The order of genes (Fig. 2) allows us to predict that a V-J recombination event involving either gene 2 or 3 (on the 8-kb *EcoRI* fragment) should result in the deletion of DNA 5' to this fragment, in this case the 14-kb *EcoRI* fragment encoding gene 1. That this did not seem to occur in the λ -producing cell line used here (see Fig. 1) provided a clue that λ -like genes were encoded on two different 14-kb *EcoRI* fragments, one that fits into this map, and a second that does not. The non-allelic nature of the second fragment was confirmed by cloning it from two individuals and then showing that these separately cloned fragments, although identical to one another, were different from the originally cloned 14-kb fragment at several restriction sites (*HindIII*, *BglII* and *CfoI*, data not shown). We have not yet established the linkage relationship of the second fragment to the six clustered genes (P.A.H. and G.F.H., unpublished data). Indeed, careful examination of the λ -producing B cell in Fig. 1 shows a reduction of ~50% in intensity of the 14-kb fragment. This is consistent with the idea that both allelic copies of the 14-kb fragment encoding gene 1 have been deleted while both allelic copies of the second 14-kb fragment have been retained. In addition, we have cloned 16-kb and 5-kb *EcoRI* fragments of germ-line DNA that exhibit weak hybridization to the human and mouse λ constant region probes. Neither of these has been linked to the major six-gene cluster (G.F.H., P.A.H. and P.L.,

unpublished results). Whether these represent pseudogenes or functional coding segments is unknown.

Sequenced genes correspond to specific human isotypes

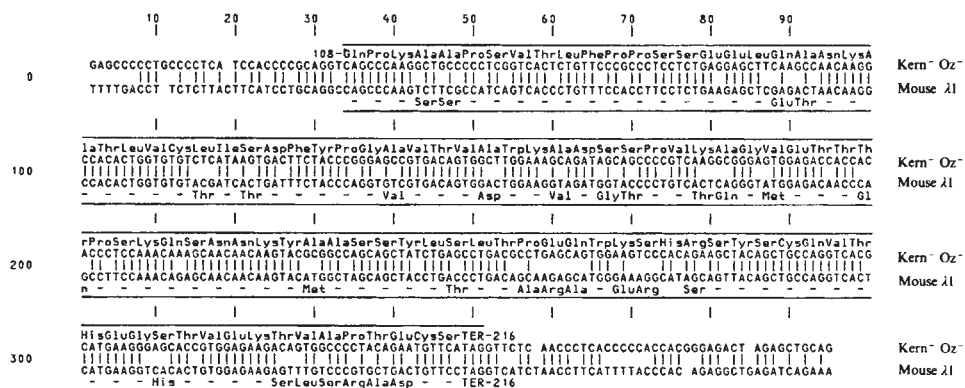
To identify the λ -like regions (indicated in Fig. 2) positively as human λ constant genes, the nucleotide sequences of the three most 5' genes were determined using the sequencing strategy

COMPARISON OF HUMAN LAMBDA CONSTANT REGION GENES (1-3)

CCC	AAG	GCC	AAC	CCC	ACG	GTC	ACT	CTG	TTC	CCG	CCC	TCC	TCT	GAG	GAG	CTC	CAA	GCC	AAC	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	3
PRO	LYS	ALA	ASN	PRO	THR	VAL	THR	LEU	PHE	PRO	PRO	SER	SER	GLU	GLU	LEU	GLN	ALA	ASN	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
AAG	GCC	ACA	CTA	GTG	TGT	CTG	ATC	AGT	GAC	TTC	TAC	CCG	GGA	GCT	GTG	ACA	GTG	GCT	TGG	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	3
LYS	ALA	THR	LEU	VAL	CYS	LEU	ILE	SER	ASP	PHE	TYR	PRO	GLY	ALA	VAL	THR	VAL	ALA	TRP	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
AAG	GCA	GAT	GGC	AGC	CCC	GTC	AAG	GGC	GGA	GTG	GAG	ACG	ACC	AAA	CCC	TCC	AAA	CAG	AGC	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	3
LYS	ALA	ASP	GLY	SER	PRO	VAL	LYS	ALA	GLY	VAL	GLU	THR	THR	LYS	PRO	SER	LYS	GLN	SER	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
AAC	AAC	AAG	TAC	GCG	GCC	AGC	AGC	TAC	CTG	AGC	CTG	ACG	CCC	GAG	CAG	TGG	AAG	TCC	CAC	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	3
ASN	ASN	LYS	TYR	ALA	ALA	SER	SER	TYR	LEU	SER	LEU	THR	PRO	GLU	GLN	TRP	LYS	SER	HIS	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
AGA	AGC	TAC	AGC	TGC	CAG	GTC	ACG	CAT	GAA	GGG	AGC	ACC	GTG	GAG	AAC	ACA	GTG	GCC	CCT	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	3
ARG	SER	TYR	SER	CYS	GLN	VAL	THR	HIS	GLU	GLY	SER	THR	VAL	GLU	LYS	THR	VAL	ALA	PRO	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
ACA	GAA	TGT	TCA	TAG																1
---	---	---	---	---																2
---	---	---	---	---																3
THR	GLU	CYS	SER	END																
-	-	-	-	-																
-	-	-	-	-																

Fig. 4 Nucleotide sequences of three human λ constant region genes. The nucleotide sequences of the coding regions of the three most 5' genes (genes 1-3; see Fig. 2) are aligned. A dash in the nucleotide sequences of genes 2 and 3 indicates that the nucleotide in that position is identical to that in gene 1. The amino acid sequences deduced are shown below the nucleotide sequences. A dash in genes 2 and 3 indicates identity in that amino acid position to the gene 1 sequence. The amino acid sequence differences (positions 112, 114, 152, 163 and 190) between the four non-allelic forms of the human λ constant region are indicated at the bottom³².

Fig. 5 Direct comparison of the nucleotide sequences of the human Kern⁻Oz⁻ λ constant gene sequence and the mouse λ 1 constant region sequence^{2,26}. The amino acid sequences deduced from the gene sequences are also shown. A dash in the mouse sequence indicates that the amino acid position is identical to that found in the human sequence. The frameshift mutation discussed in the text is at positions 253–270.



shown in Fig. 3. The nucleotide sequences corresponding to the amino acid coding regions are compared in Fig. 4 with their respective amino acid sequences.

At the protein level, at least four non-allelic forms of the human λ constant region have been identified^{16–19}. The amino acid sequence differences between these are shown at the bottom of Fig. 4. By comparing these sequences with the amino acid translation of the three genes sequenced, we identify them as Mcg, Kern⁻Oz⁻ and Kern⁻Oz⁺ (Fig. 4). Presumably, one of the other genes—4, 5 or 6—corresponds to Kern⁺Oz⁻. The remaining two genes may represent additional non-allelic forms, may be separately encoded genes that are identical in their amino acid sequences to one of the four known non-allelic forms, or they may be pseudogenes. Direct nucleotide sequence analysis will distinguish between these possibilities.

Evolutionary considerations

The extremely close homology shown by the three sequenced genes suggests that they are the result of recent gene duplications. Single nucleotide differences are scattered throughout their coding regions, the fewest being in the carboxy terminal-coding portions. For example, genes 1 and 2 are identical in the 90 3'-terminal nucleotides of their coding sequences. There are four amino acid substitutions between genes 1 and 2 and 11 third-base silent mutations. Genes 2 and 3 differ by eight silent bases but only by a single amino acid substitution.

At the amino acid level, the human Kern⁻Oz⁻ constant region is 72% homologous to the mouse λ 1 constant region, but only 61% homologous to the mouse λ 2 constant region. Therefore, it is reasonable to assume that the human genes have a more recent common ancestral antecedent with the mouse λ 1 gene. To establish the nucleotide sequence relationships between the human and mouse λ constant genes, we have compared the Kern⁻Oz⁻ nucleotide sequence with the λ 1 constant region sequence of the mouse (Fig. 5)^{2,26}. Strong homology extends throughout the coding region with single base differences being rather uniformly scattered. In the 315 nucleotides that comprise the coding block, 80 base substitutions have occurred. Thus the overall homology within the coding block at the nucleotide sequence level is 75%, which is greater than the homology observed between mouse and human κ constant sequences (68%)¹⁰. Base substitutions have occurred 19 times in the first position, 17 times in the second position and 44 times in the third position of amino acid codons, suggesting that selection occurs primarily on the basis of amino acid sequence. Curiously, one of these alterations has involved a compensating pair of single nucleotide insertion-deletion differences (positions 253 and 270; Fig. 5). A sequence characteristic of an RNA splice acceptor, PyPyXPYAg²⁷, has been conserved in each sequence immediately 5' of the coding block.

Restriction fragment length polymorphisms in the human λ gene locus

During the isolation of cloned DNA sequences that spanned the region centred about the second and third genes (Kern⁻Oz⁻ and Kern⁻Oz⁺) we observed that these clones invariably deleted

extensive segments of DNA during propagation in the bacterial host. For this reason, our library clones fell into two groups that were difficult to link (represented by clones 4, 3, 1, 5 and 2, 14, 15, 9; Fig. 2). In fact, the definite linkage of these two groups had to be established by linkage of each to the cloned 8-kb *Eco*RI fragment (Fig. 2). Thus, there seemed to be an unstable sequence associated with this region at least as recognized by the bacterial host (designated by Δ in the central 8-kb *Eco*RI fragment; see Fig. 2).

It was of interest to know whether this segment might also be unstable in man. Therefore, we prepared DNA from 15 individuals and used *in situ* hybridization to determine the arrangement of this fragment in this human population (Fig. 6). Even in such a small sample, we identified at least two, and possibly a third, arrangement of the λ constant genes within this fragment.

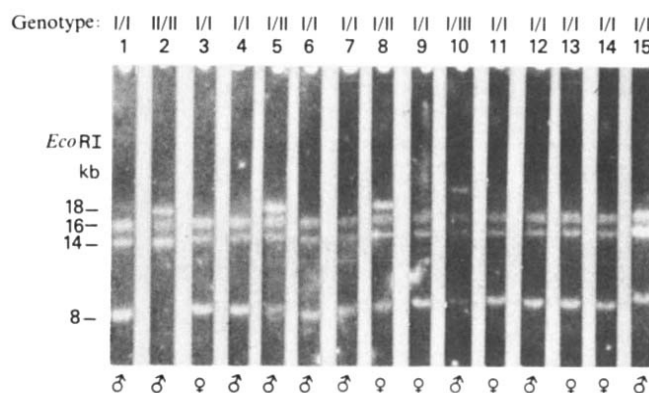


Fig. 6 Restriction fragment polymorphisms in the human C_x locus. DNA was prepared from white cells isolated from 20 ml of blood taken from 15 individuals as follows: the cells were washed twice in ice-cold phosphate-buffered saline (PBS) and pelleted at 1,000 r.p.m. for 10 min. Red cells were lysed by adding 40 ml AKC lysing buffer (0.155 M NH₄Cl, 0.01 M KCl) and left on ice for 10 min. White cells were pelleted at 1,500 r.p.m. for 10 min. DNA was extracted by resuspending the cells in 4 ml of 5% citric acid and breaking them with 10 strokes in a Dounce homogenizer (B pestle). Nuclei were pelleted by centrifugation at 2,000 r.p.m. for 5 min, then resuspended in 4 ml of 5% citric acid, layered over a 2-ml cushion of 0.88 M sucrose in 5% citric acid, and pelleted at 2,500 r.p.m. for 8 min. Nuclei were resuspended in 2 ml RSB (10 mM NaCl, 10 mM Tris pH 7.4, 25 mM EDTA) and diluted to 15 ml in RSB. Nuclei were pelleted and resuspended in 3 ml RSB. Proteinase K was added to a final concentration of 10 μ g ml⁻¹, SDS to a concentration of 1%, and the mixture incubated at 37°C for 2.5 h. The mixture was diluted to 5 ml in 0.5 M NaCl, extracted with phenol/chloroform/isoamyl alcohol (100:100:1) and centrifuged at 5,000 r.p.m. for 5 min. The aqueous phase was then extracted with ether, and precipitated by adding 2.5 vol ethanol. DNA was resuspended in 5 ml of 0.1 $\times$ SSC, digested with RNase (50 μ g ml⁻¹; 37°C for 30 min), adjusted to 0.5 M NaCl, phenol and ether extracted as before, and ethanol-precipitated. The DNA was resuspended in 10 mM Tris pH 7.8, 5 mM NaCl, 0.5 mM EDTA and stored at 4°C. Twenty ml of blood yielded ~400 μ g of DNA. DNA (10 μ g) was digested with *Eco*RI restriction endonuclease, size fractionated by agarose gel electrophoresis, transferred to nitrocellulose, hybridized to a human λ constant region probe (combined 1.2-kb *Bam*HI-*Eco*RI fragment containing the Kern⁻Oz⁻ C_x gene and 0.8-kb *Bam*HI-*Hind*III fragment containing the Mcg C_x gene (see map, Fig. 2)), and autoradiographed. The lengths of the DNA fragments are indicated alongside. Three gene configurations were noted (types I, II and III) and are described in the text.

These polymorphisms represent differences in the lengths of the restriction fragments on which the λ constant genes $\text{Ke}^-\text{Oz}^-(2)$ and $\text{Ke}^-\text{Oz}^+(3)$ reside. Case 1 is the most common genotype (type I), characterized by 8-, 14- and 16-kb *EcoRI* restriction fragments. Case 2 has a different genotype in which an 18-kb fragment replaces the 8-kb fragment. Indeed, case 2 is homozygous for the 18-kb fragment as no 8-kb fragment is present in this individual. Case 5 represents an individual who is heterozygous for these alleles (type I/II). The allelism of these fragments has been further established by familial studies (P.A.H., G.F.H. and P.L., unpublished results). There is a third polymorphism (case 10, type III) associated with a 21-kb *EcoRI* fragment. As we have only one example of this type and because it retains at least one 8-kb fragment, we cannot be certain that the 21- and 8-kb fragments are allelic. The band intensity (Fig. 6) suggests, however, that this is the case.

Obviously these polymorphisms could arise either by point mutations or by rearrangements of large fragments of genomic DNA. The former mechanism seems less likely because it would require the alteration of several *EcoRI* sites to create an 18-kb (or larger) *EcoRI* fragment from the 8-kb fragment seen in type I (see map, Fig. 2). Rather, the labile behaviour of this fragment in *E. coli* suggests that it contains a 'hot spot' that readily deletes a segment of DNA. This, together with the frequent evolutionary rearrangement of DNA in higher organisms^{24,28}, makes the rearrangement model a more attractive one.

Received 7 August; accepted 2 October 1981.

1. Brack, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
2. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15**, 1133–1144 (1978).
3. Seidman, J. G. & Leder, P. *Nature* **276**, 790–795 (1978).
4. Lenhard-Schuller, R., Hohn, B., Brack, C., Hiram, M. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4709–4713 (1978).
5. Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280**, 370–375 (1979).
6. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11–17 (1978).
7. Valbuena, O., Marcu, K. B., Weigert, M. & Perry, R. P. *Nature* **276**, 780–784 (1978).
8. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450–3454.
9. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
10. Weigert, M. *et al. Nature* **283**, 497–499 (1980).
11. Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V. & Leder, P. *Cell* **22**, 197–207 (1980).
12. Miller, J., Bothwell, A. & Storb, U. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3829–3833 (1981).
13. Blomberg, B., Trautnecker, A., Eisen, H. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3765–3769 (1981).
14. Hieter, P. A., Korsmeyer, S. J., Waldmann, T. & Leder, P. *Nature* **290**, 368–372 (1981).
15. Korsmeyer, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).

Multiple J_λ - C_λ regions and J_λ repertoire

We have already noted that there are several important differences in the germ-line arrangements of κ and λ genes, a main one (noted in the mouse) being that while there are at least four active J regions available for V-J recombination for the single κ region gene^{4,5,8,9}, there is only one J for each of the mouse λ genes so far characterized^{12,13}. *A priori*, this would limit the diversity that could be generated through V-J recombination in the λ system if there was only one λ constant region. The complexity of the human and mouse C_λ loci suggests a mechanism that may compensate for this deficiency. Instead of creating a strip of J regions in front of a single constant region (as seen in the κ locus), the entire λ region in the mouse has been amplified to create four J regions, each associated with its own C region. The net effect is to make four J regions available in both the κ and λ systems. In the human λ gene system, the options for using each constant region may similarly depend on V-J recombination, but the complex organization of the locus allows us to consider subsequent recombination steps (akin to heavy chain switching) or even alternative modes of RNA processing as mechanisms for activating other J_λ - C_λ segments. The complete characterization of this locus and precise identification of its J regions will allow us to distinguish among these possibilities.

We thank Terri Broderick for help in preparing the manuscript and Barbara Norman and Marion Nan for assistance.

16. Ein, P. *Proc. natn. Acad. Sci. U.S.A.* **60**, 982–985 (1968).
17. Gibson, D., Levannon, M. & Smithies, O. *Biochemistry* **10**, 3114–3122 (1971).
18. Fett, J. & Deutsch, H. F. *Immunogenetics* **12**, 643–652 (1975).
19. Solomon, A. *Immunogenetics* **5**, 525–533 (1977).
20. Fett, J. W. & Deutsch, H. F. *Immunogenetics* **13**, 149–155 (1976).
21. Lieu, T. S., Deutsch, H. F. & Tischendorf, T. W. *Immunogenetics* **24**, 429–433 (1977).
22. Lopez de Castro, J. A., Chiu, Y. Y. H. & Poljak, R. J. *Biochemistry* **17**, 1718–1723 (1978).
23. Maniatis, T. *et al. Cell* **15**, 687–702 (1978).
24. Wymann, A. R. & White, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6754–6758 (1980).
25. Polsky, F., Edgell, M. H., Seidman, J. G. & Leder, P. *Analyt. Biochem.* **87**, 397–410 (1978).
26. Bothwell, A. L. M. *et al. Nature* **290**, 65–67 (1981).
27. Seif, I., Khoury, G. & Dhar, R. *Nucleic Acids Res.* **6**, 3387–3389 (1979).
28. Konkel, D., Maizel, J. V. Jr & Leder, P. *Cell* **18**, 865–873 (1979).
29. Southern, E. J. *molec. Biol.* **98**, 503–517 (1975).
30. Rigby, P. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
31. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
32. Hood, L. E., Weissman, I. L. & Wood, W. B. in *Immunology*, 234 (Benjamin/Cummings, Menlo Park, California, 1978).

LETTERS TO NATURE

Radio jet of 3C273

R. G. Conway, R. J. Davis & A. R. Foley

University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

T. P. Ray

University of Manchester, Department of Astronomy, Manchester M13 9PL, UK

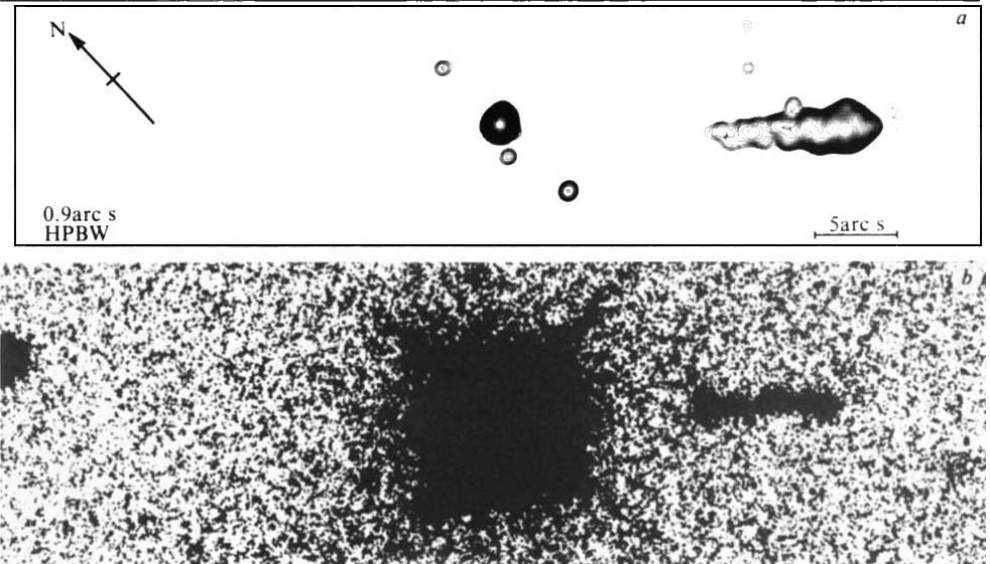
Most radio sources are two-sided, but a minority appear one-sided, 3C273 being the first-known and brightest example. There is no agreement^{1–4} on whether such sources are intrinsically one-sided, or are normal double sources, one half of which is hidden by Doppler effects. We report here new radio observations at 408 MHz of 3C273 which show that the brightness of the postulated counter-jet is $<1/100$ of the brightness of the visible jet. If this ratio is due to Doppler beaming, the source must be seen almost end-on, and the whole jet must be moving at a quasi-relativistic speed ($>0.7c$) into an ambient medium with number density $<0.6 \text{ m}^{-3}$. Because several arguments suggest that such a density is implausibly low, the jet of 3C273 cannot be identified with the radio lobe of a normal double source. If the emitting regions are moving slowly the ejection from the nucleus is certainly to one side only. An alternative possibility is that the jet is moving relativistically and behaves as one of the fast beams in the beam-model of Blandford and Rees⁵.

New radio observations of 3C273 at 408 MHz ($\lambda 0.735 \text{ m}$) made with the MTRLI^{6,7} at Jodrell Bank are shown in Fig. 1a, along with a reproduction to the same scale (Fig. 1b) of an optical photograph⁸, taken in the blue-green. (The optical QSO is 13th magnitude, and has a redshift of 0.158.) In neither picture is there any evidence of a counter-jet, and from the radio map a limit of $<1:100$ may be put on the brightness ratio of the two sides. The lack of significant emission either at radio or optical wavelengths in the region out to 12 arc s. from the quasar is suggestive of an invisible beam⁵ of material with ordered relativistic motion or of photons carrying energy to the outer component. However, we cannot rule out an alternative explanation of the 'gap', namely that the output power of the quasar has decreased significantly during the past $\sim 10^5$ yr.

The radio emission in Fig. 1a shows 'wiggles' or oscillations from side to side of the central ridge-line, a phenomenon first suggested by Conway and Stannard⁹, the general features of whose map we confirm. The wiggles are not accompanied by the oscillations in brightness which would be expected if they were due to precession of the source axis. We suggest that they are more probably hydrodynamic instabilities¹⁰ of the jet as it traverses the intergalactic medium.

Table 1 lists some observed parameters of the radio emission, which yield physical quantities that can be compared with predictions from different source models, in particular the model with a relativistically moving counterjet, which is hidden by Doppler beaming. We have calculated the magnetic field and

Fig. 1 *a*, The MTRLI map of 3C273 at 408 MHz ($\lambda 0.735$ m) tilted so that position angle 223° is horizontal. The contour levels are logarithmic, with three contours to a factor 2 in brightness, the lowest contour being 0.12 Jy per beam. The HPBW of the beam is shown as a circle 0.9 arc s in diameter in the bottom left corner. The 3% sidelobes north and south of the central quasar are artefacts due to calibration uncertainties, and may be ignored. *b*, Photograph by Arp⁸ (with the same scale as *a*). Seeing is about 1 arc s. Neither *a* nor *b* show any evidence for a counter-jet, and display a gap in emission out to 12 arc s. from the central nucleus. The optical jet appears to have an approximately constant brightness along its length with a possible knot of emission nearest the quasar, whereas the radio brightness increases outwards by at least 100:1. The radio emission also displays 'wiggles' along the length of the tail. If $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, 1 arc s equals 1.85 kpc.



energy density in the head from the spectral index and brightness using standard synchrotron theory¹¹. If there is equipartition of energy between the magnetic field and relativistic particles (the latter term divided equally between ions and electrons) and if the bulk motion is zero, the field B is 29 nT and the energy density u_{obs} is 680 pJ m^{-3} . Multifrequency measurements of the radio polarization of the jet (R.J.D., unpublished data) show that the emission depolarizes at $\lambda \approx 0.73$ m. If this depolarization is due to internal Faraday rotation by thermal matter²⁰, then the electron density n_e within the jet is 11 m^{-3} . (This value is unusually low. If the depolarization is external, n_e must be lower still.) The sound speed, equal to $(\frac{1}{3} \text{ energy density/matter density})^{1/2}$, is calculated to be $0.29c$. Such a value means that the jet must be confined sideways by some means, because if it were in free expansion, the opening angle of the jet would be much larger than is observed. This conclusion is not altered if the motion of the jet is relativistic.

VLBI measurements¹² have shown the presence of 'superluminal' proper motions in the nucleus of the quasar. According to the ballistic model (see ref. 2) the most probable value of θ , the angle to the line of sight, is related to the true velocity $\beta_n c$ by:

$$\beta_n = \cos \theta \quad (1)$$

This condition also results in the maximum enhancement of proper motion². If equation (1) is satisfied, and if Hubble's constant H_0 is $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (a value indicated by recent work^{13,14}), then $\theta \approx 11^\circ$, $\beta_n = 0.983$ and the Lorentz factor $\gamma_n = 5.4$. We assume that the jet also lies at $\theta = 11^\circ$, in which case the length D from the nucleus to the outer head is 200 kpc. (If the 'light-echo' model is adopted¹⁵, then $\theta = 22^\circ$ and $D = 100$ kpc.)

We assume that the forward speed of the jet is $\beta c \leq \beta_n c$, with Lorentz factor $\gamma \leq \gamma_n$. The blueshift factor is

$$\eta = \frac{\nu_{\text{observed}}}{\nu_{\text{emitted}}} = \frac{1}{\gamma(1 - \beta \cos \theta)} \quad (2)$$

It can be shown² that the energy density in a moving jet needed to produce the observed brightness is

$$u_{\text{rest}} = u_{\text{obs}} \eta^{-(2+\alpha)/4/7} \quad (3)$$

The forward motion of the jet will be opposed by a ram-pressure force proportional to the number density n_m of the medium in front of the head. For relativistic speeds the ram-pressure, referred to the rest frame of the head, is

$$P_{\text{ram}} = n_m \mu (\gamma \beta c)^2 \quad (4)$$

where μ is the average mass per particle in the ambient medium. The motion will be decelerated unless $P_{\text{ram}} \approx \frac{1}{3} u_{\text{rest}}$. From these equations, n_m may be found from the measured brightness as a function of the forward speed βc :

$$n_m = 2.2 (\beta \gamma)^{-2} \eta^{-1.54} \text{ m}^{-2} \quad (5)$$

This relation is shown graphically in Fig. 2.

If the absence of the counter-jet is due to Doppler beaming, then the speed required would be $\beta > 0.7$ ($\gamma > 1.4$), and the ambient density $n_m < 0.6 \text{ m}^{-3}$. Such a low density is implausible because: (1) it is even lower than the universal density, estimated to be 10% of the closure density from quasar statistics¹⁶; (2) 3C273 is a member of a group of galaxies¹⁷ where the expected density would be intermediate between the universal value and the value ($\sim 10^{+3} \text{ m}^{-3}$) found in large clusters¹⁸; (3) the formation of a group of galaxies and a massive quasar presents difficulties in a region with a very low density; (4) the sideways confinement of the jet requires a density several orders of magnitude higher than 0.6 m^{-3} .

The ram pressure argument may be used in reverse. If n_m lies between the closure density ($\approx 20 \text{ m}^{-3}$) and the upper limit of $\sim 5,000 \text{ m}^{-3}$ set by X-ray observations (R. Willingale, personal communication), then the forward speed is in the range $0.02 < \beta < 0.3$. This value refers to the speed of the bow-shock in front of the head. In conventional radio source models much of the emission is from a 'working surface' moving at a speed close to that of the bow-shock. If this is the case in 3C273, a receding counter-jet with $\beta < 0.3$ would be detectable well above our sensitivity limit, and we conclude that the ejection must be to one side only.

A source model with one-sided ejection (model A) still has unexplained features. Unless the temperature of the ambient gas exceeds 10^8 K , hydrostatic gas pressure will not confine the jet sideways, and some other mechanism must be cited to

Table 1 Observed parameters of 3C273

Projected separation between nucleus and head of jet	40 kpc
Size of head	$2.1 \text{ kpc} \times 1.1 \text{ kpc}$
Apparent opening angle of jet	0.03 rad
Spectral index $\alpha (S \propto \nu^{-\alpha})$	0.7
Depolarization wavelength $\lambda^{1/2}$ of head ²⁰	0.7 m
Brightness temperature of head	$2.4 \times 10^8 \text{ K}$

achieve confinement. A further consequence of model A comes from the following statistical argument (see Fig. 3 and ref. 4). If one-sided sources commonly consist of relativistically moving cores (emitting radiation beamed into a solid angle $\sim \gamma^{-2}$) together with slow-moving outer lobes emitting unbeamed radiation, then for every 'core-dominated' source, there should be $\sim \gamma^2$ sources with a single lobe and little or no core. The total number of one-sided core-dominated sources exceeds 50 (I. W. A. Browne, personal communication) but no single example has yet been found of a one-sided source with no core. Hence, the motion in the core of a one-sided source cannot be at an appreciably higher γ than the motion in the outer lobe. If this argument is accepted, the superluminal velocity observed in the core of 3C273 cannot be interpreted as a ballistic motion, and must instead be a light-echo or similar effect.

We may avoid these conclusions if we assume that in 3C273 the radio radiation is not emitted from slow-moving regions such as the 'working surface', but instead is emitted directly from a fast beam, such as that proposed in the beam-model by Blandford and Rees⁵. We have explored a model along these lines (model B) in which the ejection is relativistic and may (but need not) be two-sided. The superluminal proper motion in the nucleus corresponds to true ballistic motion with $\gamma \sim 5$, and this velocity persists unchanged out to distances of 200 kpc (20 arc s) from the nucleus. Because of the high Doppler beaming factor (~ 100) the radio emission from the fast beam will dominate over that from slow-moving material. However, the absence of a detectable counter-jet implies that the rest-frame luminosity of the fast beam is at least as great as that of the working surface, which would seem unlikely in view of the greatly increased turbulence at the working surface⁵. This difficulty is avoided if the ejection is one-sided even in model B.

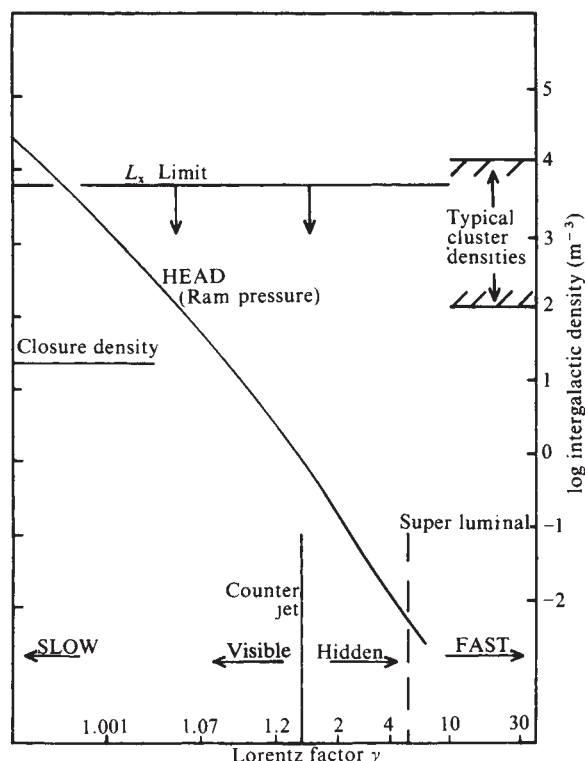


Fig. 2 The density of the intergalactic medium which will confine the head of 3C273 by ram-pressure as a function of forward speed, expressed as the Lorentz factor, γ , of the head. The actual function used is $n_m = 2.2 \beta^{-2} \gamma^{-2} \eta^{-1.54} \text{ m}^{-3}$. The L_x limit is provided by unpublished results from the Einstein satellite (R. Willingale, personal communication) and assumes a temperature of 10^8 K for the intergalactic medium. The speed at which an intrinsically equal counter-jet will be hidden by the Doppler effect is indicated, as is the value corresponding to the superluminal expansion¹².

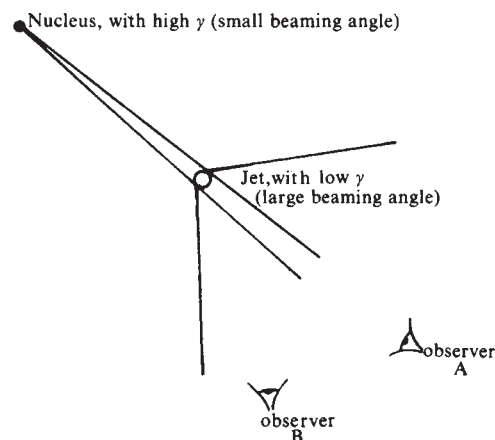


Fig. 3 Model for a radio source with a high γ in the nucleus and a low γ in the jet. Observer A sees strong radiation from both nucleus and jet. However a more typical observer B would see strong radiation from the jet but little radiation from the nucleus.

If the radio emission is from the fast beam directly, the beam must consist of relativistic particles, rather than photons. The thermal matter responsible for Faraday depolarization must not be co-moving, and hence cannot lie within the jet. It must form an irregular foreground screen, very possibly in the 'cocoon' region surrounding the beam. The front bow-shock moves slowly, and the ram pressure condition may be satisfied by the same values as in model A ($0.02 < \beta < 0.3$, $5000 > n_m > 20 \text{ m}^{-3}$). There may be a difficulty with the power requirements in model B: to supply the fast-moving beam continuously requires at least $5 \times 10^{38} \text{ W}$, which is 1% of the power output of the central quasar¹⁹ and 100 times the emitted radio power.

The difficulties associated with these energy problems which are not unique to 3C273 have been discussed by Rees³. In the case of 3C273 the power flowing down the jet in the form of random particle motions is at least a factor $100^{4/7}$ greater at the head than at the innermost end of the jet, which suggests that (in model B) the jet of 3C273 is not a continuous flow system.

We conclude that if the jet of 3C273 corresponds to the radio lobe of a normal radio source, then the emission is virtually unbeamed, and the ejection of material from the nucleus must be to one side only. Alternatively, the jet may be a visible example of a fast beam, such as that proposed in the beam model⁵, in which case the ejection may be two-sided, as Doppler beaming will suffice to hide the receding counterpart. Other factors have to be invoked, however, to explain the absence of unbeamed radiation from the radio lobe which such a fast beam would normally produce.

We thank Professor H. C. Arp for use of his recent photograph of 3C273. We also thank I. W. A. Browne and P. Birch for detailed discussions on relativistic beams. A.R.F. acknowledges an SRC research studentship.

Received 14 September; accepted 19 October 1981.

1. Scheuer, P. A. G. & Readhead, A. C. S. *Nature* **277**, 182–185 (1979).
2. Blandford, R. D. & Königl, A. *Astrophys. J.* **232**, 34–48 (1979).
3. Rees, M. J. *IAU Symp.* No. 94, 139–164 (1981).
4. Browne, I. W. A. *et al. Mon. Not. R. astr. Soc.* (in the press).
5. Blandford, R. D. & Rees, M. J. *Mon. Not. R. astr. Soc.* **169**, 395–415 (1974).
6. Davies, J. G., Anderson, B. & Morison, I. *Nature* **288**, 64–66 (1980).
7. Cornwell, T. J. & Wilkinson, P. N. *Mon. Not. R. astr. Soc.* **196**, 1067–1086 (1981).
8. Arp, H. *Proc. ESO/ESA Workshop SP-162*, 53–61 (1981).
9. Conway, R. G. & Stannard, D. *Nature* **255**, 310–312 (1975).
10. Ray, T. P. *Mon. Not. R. astr. Soc.* **196**, 195–207 (1981).
11. Pacholczyk, A. G. *Radio Astrophysics* (W. H. Freeman, San Francisco 1970).
12. Pearson, T. J. *et al. Nature* **290**, 365–368 (1981).
13. Aaronson, M. *et al. Astrophys. J.* **239**, 12–37 (1980).
14. de Vaucouleurs, G. & Bollinger, G. *Astrophys. J.* **233**, 433–452 (1979).
15. Lynden-Bell, D. *Nature* **270**, 396–399 (1977).
16. Sherman, R. D. *Astrophys. J.* **246**, 365–368 (1981).
17. Stockton, A. *IAU Symp.* **92**, 89–97 (1980).
18. Silk, J. A. *Rev. Astr. Astrophys.* **11**, 269–308 (1973).
19. Ulrich, M.-H. *Space Sci. Rev.* **28**, 89–104 (1981).
20. Conway, R. G. *et al. Mon. Not. R. astr. Soc.* **168**, 137–162 (1974).

Halo around the Crab Nebula

Paul Murdin

Royal Greenwich Observatory, Herstmonceux Castle, Hailsham,
East Sussex BN27 1RP, UK

David H. Clark

Rutherford Appleton Laboratory, Chilton, Didcot,
Oxfordshire OX11 0QX, UK

It has been argued¹ that the supernova of AD 1054 which left the Crab Nebula supernova remnant was of Type II. Such supernovae occur in the spiral arms of galaxies and are believed to have initial masses of $>4M_{\odot}$. The helium overabundance detected in the Crab Nebula filaments is compatible² with an initial stellar core mass of $2\text{--}4M_{\odot}$ corresponding to a total precursor mass of $10\text{--}15M_{\odot}$. The Crab Nebula as usually depicted consists of filaments of total mass $2\text{--}3M_{\odot}$ and a pulsar of $1.4M_{\odot}$. If the precursor of SN1054 was heavier than $\sim 4M_{\odot}$, its outer envelope could be detectable¹ as an extended halo around the Crab Nebula. We have detected such an outer halo, whose size implies that the progenitor of SN1054 was indeed a massive star.

The observational material was a deep IIIaF plate obtained on 28 December 1976 from the 1.2-m UK Schmidt Telescope specifically to search for a halo around the Crab, and partly analysed but displaced from our attention by other interesting observations in June 1978. The exposure was through an interference filter with a $150\text{ \AA}$ bandpass centred on $H\alpha$, and was calibrated with a step wedge illuminated with broadband red light. The calibration curve had to be extrapolated from densities ~ 0.3 to the sky density ~ 0.2 , very near the toe of the curve, and our estimate of the intensity of the halo must be treated with caution. An area 40×40 times arc min around the nebula was scanned on the PDS microdensitometer at the Anglo-Australian Observatory and calibrated by the step wedge exposed on the plate. The point-spread-function (p.s.f.) of the instrumentation is shown by images of stars and has a circular disk $\sim 1.5\text{ mm}$ (1.7 arc min) in diameter (caused by multiple reflections). Several star images showing this disk arising from a central density comparable with that deep within the nebula were also scanned in the PDS machine to determine the p.s.f. beyond the disk. It drops to zero ($<1\%$ of night sky brightness) at 2 arc min radius from the stars.

The elliptical image of the Crab Nebula has a bright central area of 7 arc min major axis by 4 arc min minor axis, in accord with the dimensions derived by van den Bergh³. Beyond this abrupt boundary is a halo whose brightness falls to the night sky background at a distance from the centre of the nebula of 6 arc min along the minor axis direction (Fig. 1) and 14 arc min along the major axis direction, well beyond the tail of the p.s.f. The dimensions of the halo are thus three to four times that of the nebula.

The surface brightness of the halo is $\sim 2\%$ of the night sky intensity. At Siding Spring, Bessell⁴ measured the R -band sky brightness at $21.9\text{ mag arc s}^{-2}$. The night sky brightness is a variable reference, depending on season and solar cycle, but through the filter the Crab halo brightness is $\sim 2 \times 10^{-7}\text{ erg s}^{-1}\text{ cm}^{-2}\text{ sr}^{-1}$. This corresponds⁵ to an emission measure for the $H\alpha$ intensity of $\sim 7\text{ pc cm}^{-6}$, with an uncertainty of a factor of two, corresponding to errors in the plate calibration and reference source.

A thermal soft X-ray enhancement peripheral to the nebula has been detected⁶ from a lunar occultation observation of the Crab. Imaging soft X-ray observations from the Einstein Observatory have failed to confirm the reality of the halo because of the problems of scattering of the intense nebular

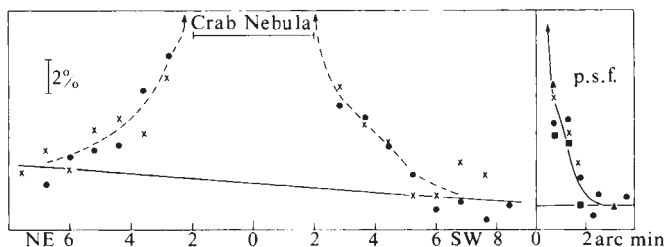


Fig. 1 Two surface brightness profiles (● and ×) along the minor axis of the Crab Nebula. The profiles are from parallel strips 1.6 arc min wide lying either side of the minor axis. The Crab Nebula has a bright central area 4.0 arc min in minor axis, with a halo extending 4 arc min beyond this. Also shown is the point-spread function of the instrument and measuring technique. The night-sky background under the Crab Nebula shows a slope up to the Milky Way in the north-east.

emission from the telescope mirror; modelling of this scattering may yet reveal a residual halo (F. Seward, personal communication). An optical halo extending 20% further from the centre than the abrupt edge of the filaments has been detected by Scargle⁷ in the continuum bandpass $530\text{--}590\text{ nm}$, and attributed to leaking electrons and magnetic fields, yielding extended synchrotron emission. As the halo we have detected through the $H\alpha$ filter extends further and appears to be stronger we attribute it to $H\alpha$ emission, and Scargle's halo we suppose to be weak line emission.

The origins of nebula and halo are proposed to be the following¹. The core collapse deposited $\sim 2 \times 10^{51}\text{ erg}$ in the helium-rich outer core, which, expanding outwards, was decelerated, thus setting up a high-pressure region. The consequent reverse shock wave drove the core material back to the centre. As the original deceleration was Rayleigh-Taylor unstable, some mixing of outer-core and envelope is expected—only $0.5M_{\odot}$ of mixed envelope material is required to provide the observed nebula abundances¹. The bulk of the supernova energy was transferred to the outer envelope of the presupernova star. On this model the observed nebula is the outer core-envelope mixture, whose present kinetic energy has been derived almost entirely from the pulsar, and the postulated halo is high-velocity low-density material from the envelope of the precursor. A consequence of such an interpretation is that the near coincidence between the supernova in history and the proper motion age for the nebula is merely coincidence.

An alternative explanation for the halo is that it represents light scattered off dust grains surrounding the Crab. These may have been produced by the precursor (an alternative indication that it was massive) or be swept-up interstellar material (J. C. Wheeler, personal communication) even at the low ambient interstellar density near the Crab ($z \sim 200\text{ pc}$ above the galactic plane).

The size and brightness of the observed optical halo compare well with the properties of the outer shell of the Crab predicted by Chevalier¹. He calculated that a shell moving at $5,000\text{ km s}^{-1}$ would have a size R four times the radius of the Crab. We observe a halo $3\text{--}4$ times the Crab radius. If the shell was of mass $M = 10M_{\odot}$, mean density 1 atom cm^{-3} , it would be fully ionized by the Crab's synchrotron radiation and have an emission measure of 5 pc cm^{-6} . We observe $E \sim 7\text{ pc cm}^{-6}$. Thus the halo we observe may be slightly more compact and brighter than the prediction. The mass in the shell is most sensitive to the extent of the halo: M scales as $E^{1/2}R^{5/2}$, yielding $M \sim 8M_{\odot}$, so Chevalier's prediction is satisfactory. Adding filaments and pulsar, the precursor mass was about $12M_{\odot}$.

The above calculation is based on the hypothesis of a uniform shell density. M is overestimated if the shell is lumpy or structured. The lack of granularity and the lack of peripheral brightening in the shell we observe are surprising but a massive progenitor would be expected to experience significant mass-loss before it went supernova and a substantial density gradient surrounding the expanding envelope might explain the observed

brightness distribution. In view of uncertainties in the model, in the reference intensity of the night sky and in the calibration of the plate, our estimate of the Crab precursor's mass of $12 M_{\odot}$ is indicative only.

Observations of the halo, such as photoelectric confirmation of its extent, distribution and brightness, spectral confirmation of its emission spectrum and detection of its anticipated large velocity dispersion, and polarization measurements are needed to elucidate its nature.

We thank Peter Krug for measuring the plate, which was obtained by Tim Hawarden.

Received 27 August; accepted 27 October 1981.

1. Chevalier, R. *Supernovae* (ed. Schramm, D.) 53 (Reidel, Dordrecht 1977).
2. Arnett, W. D. *Astrophys. J.* **195**, 727 (1975).
3. van den Bergh, S. *Astrophys. J. Lett.* **160**, L27 (1970).
4. Bessell, M. *Publ. astr. Soc. Pacif.* **91**, 603 (1979).
5. Allen, C. W. *Astrophysical Quantities*, 3rd edn (Athlone, London, 1974).
6. Toor, A., Palmieri, T. M. & Seward F. D. *Astrophys. J.* **207**, 96 (1976).
7. Scargle, J. D. *Publ. astr. Soc. Pacif.* **82**, 388 (1970).

Spin directions of interfering beams in quantum interferometry

A. Zeilinger*

Atominstut der Österreichischen Universitäten, A-1020 Wien, Austria

It is shown here that, in a fermion interferometer experiment, the notion of the relative orientation of the interfering beams can be given a sensible meaning—information may be extracted about the spin directions without destroying the interference pattern. This is done on a gedanken-experiment level by introducing into, say, a neutron interferometer Stern–Gerlach magnets with detectors placed into their respective ‘down’-spin beam paths. Similar considerations apply to experiments where spinor behaviour has been demonstrated in systems with higher spin.

In interferometry experiments with non-zero spin particles interesting effects arise due to the wave function (or probability amplitude) of the interfering beams being nonscalar. One effect concerns the sign change of a spinor wave function under 2π -rotations^{1,2}. This was verified by neutron interferometry experiments^{3–7}, in which an incident beam was split coherently into two wave trains one of which was passed through a static magnetic field where it experienced Larmor precession and hence spin rotation. The spinor property was then found as a variation of the intensity of the recombined beams with a fringe period of multiples of 4π of the rotation angles. The values of the rotation angles were calculated from the magnetic field strengths using the known value of the neutron magnetic moment.

The interpretation of this type of experiment has been widely debated^{8–12}. Moore⁸, in discussing Bernstein's proposal² to use Larmor precession to implement the desired rotation, claims that in this experiment it would not be possible to observe directly that anything is actually rotating.

Byrne⁹ has explicitly demonstrated the differences between spinor and vector wave interferometry. He showed that, in a spinor experiment, it is not possible to observe simultaneously the interference pattern and to obtain information about the relative spin directions of the constituent beams from measurements on the recombined beam. Byrne concludes that the notion of relative rotation ceases to have a meaning as it corresponds to nothing which is measurable. Thus he dismisses the interpretation that the experiments have provided a direct observation of the sign reversal of a spinor wave function subjected to a 2π rotation as unsatisfactory because, for any

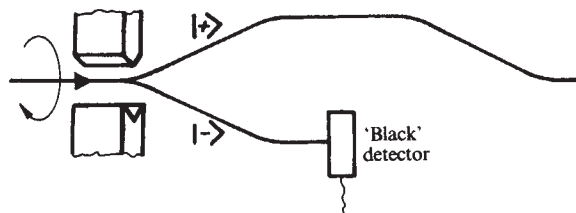


Fig. 1 Modified Stern–Gerlach arrangement with a black detector in one beam path.

fermion the relative rotation of the spins in the two beams and the interference pattern are mutually incompatible observables.

However, it is possible to conceive of experimental schemes where information may be extracted from the beams interfering about their spin directions without destroying the interference pattern. In our experiments unpolarized incident neutrons were used for intensity reasons only. We consider here polarized incident neutrons that are able to trace the behaviour of the interfering spin states.

We look for a measurement procedure which changes neither the amplitudes nor the relative phase of these beams—a measurement without wave packet reduction. In quantum mechanics, such a measurement is possible if the quantum system under investigation is in an eigenstate of the observable defined by the measuring apparatus. For spin measurement, we consider for example, a Stern–Gerlach experimental arrangement in which we place into one of its separated beam paths a black (100% absorbing) detector (Fig. 1). To restore the beams to their initial paths, we place behind the first Stern–Gerlach an inverted second one. If the incoming beam is in the eigenstate corresponding to the detector-free path no event will be registered in the detector. The restriction that the amplitude at the detector never vanishes exactly due to the nonlocality of wave packets, can be reduced by increasing the separation of the beams. Similar considerations apply to the non-existence of exactly 100% absorbing detectors. Thus, the result that we did not register events in the detector together with the knowledge that particles have passed through the Stern–Gerlach gives us information about the spin observable of the incident beam.

In the more general case where a spin rotation device is placed in front of the Stern–Gerlach, the particles are no longer in an eigenstate of the Stern–Gerlach. Therefore, events will be registered in the detector. For simplicity we assume the incident beam to be polarized in a direction normal to its propagation direction and the magnetic field to rotate the spin around that propagation direction. The spin direction of the beam can then be found by rotating the Stern–Gerlach around the neutron propagation direction to that angular position where the counting rate in the detector vanishes. This spin direction is a macroscopically observable quantity—the angular position of the Stern–Gerlach. Many other geometrical arrangements can be envisaged on a gedanken-experiment level because for slow particles such as thermal neutrons the spin can be made to point in any direction relative to the propagation direction.

Formally we may describe our arrangement by a projection operator of the form $P = |s\rangle\langle s|$, where $|s\rangle$ is the +eigenstate of the Stern–Gerlach. If that operator acts on the incident state $|s\rangle$, that state is reproduced.

In an interferometer experiment of the type shown in Fig. 2 an incoming wave is split into two partial waves by a semireflecting mirror. These waves follow two distinct paths within the interferometer. Furthermore, in one beam path a magnetic field is arranged to rotate the spins. To observe the spin directions of these interfering beams we place a modified Stern–Gerlach arrangement into each interferometer beam path. In both arrangements we again place a detector into one of its beam paths, say, the ‘down’ spin beam path, leaving the ‘up’ beam path through each Stern–Gerlach free. If we start the experiment with arbitrary angular settings of the analysing direction of the Stern–Gerlach magnets we will again register particles in

* Present address: Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

the detectors. The count rate in each detector will be lower in the interference experiment because the amplitude of each separated interferometer beam is lower than the amplitude of the incoming wave. But the fact that particles are registered means that the corresponding partial wave is not in the 'up' eigenstate of the Stern-Gerlach in that partial beam.

Both Stern-Gerlachs are now rotated independently of each other until we arrive at a position where neither detector registers a single event. At that point, the amplitude of the corresponding partial wave in the 'down' eigenstate of the Stern-Gerlach vanishes. However, we can still observe the interference pattern of the recombined beams, because our Stern-Gerlach arrangement does not enable us to determine which path is followed by the particle in the interferometer. Hence, both the null-effect spin measurements and the interference fringe observation allow us to conclude that the particles are in a coherent superposition of the two 'up' states defined by the two different Stern-Gerlachs. Or $|\psi\rangle = |S_1\rangle + |S_2\rangle$, where $|S_1\rangle$ and $|S_2\rangle$ are the 'up' eigenstates of the Stern-Gerlachs in the beam paths 1 and 2, respectively.

As the analysing directions of the Stern-Gerlachs are macroscopically observable quantities we conclude that the relative spin directions of the interfering beams are operationally meaningful quantities. This also applies to the relative rotation of the interfering beams because, in principle, we can trace the effect of the spin rotation device in arbitrarily small steps. We still have to consider whether the introduction of our spin-measurement devices may destroy the relative phase of the interfering beams.

Nevertheless, when considering—on the gedanken-experiment level—the use of the Stern-Gerlach magnets the inhomogeneous magnetic fields may lead to a destruction of the coherence of the interfering beams¹³. But other spin measurements can be envisaged where this is not the case. For thermal neutrons, diffraction at perfect crystals in external homogeneous magnetic fields¹⁴ and refraction at wedge-shaped magnetic fields¹⁵ both lead to a separation of the spin states while retaining their coherence properties. These schemes also offer the advantage that in contrast to a gedanken-experiment they are actually realizable experimentally along the lines of the existing technology of perfect crystal neutron optics.

At about the same time that the neutron interferometer experiments were performed it was realized¹⁶ that the spinor behaviour can also be observed in a spin 1 system. This exploits the fact that the wave function of a two-state system—except for the photon and similar degenerate cases—changes sign under a 2π rotation. Thus, if in a three-state system the populations of two levels are inverted twice by properly phased electromagnetic high frequency pulses, their state function acquires the -1 phase factor. That phase factor may be revealed by observing interferences between one or both of these two states and the third state. There is also experimental evidence¹⁶ for that

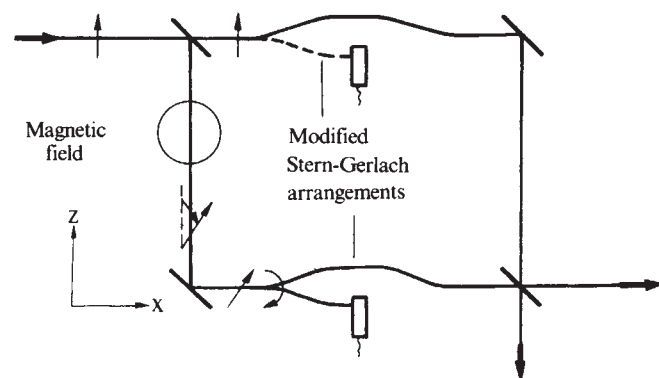


Fig. 2 Fermion interferometer gedanken-experiment with modified Stern-Gerlach arrangements in both beam paths. The magnetic field produces a spin rotation around the x -direction. A corresponding rotation of the lower Stern-Gerlach ensures unattenuated passage.

behaviour based on radio-frequency transitions between hyperfine energy levels in TIF. The same effect can be demonstrated experimentally by exploiting NMR transitions¹⁷⁻¹⁹. There, systems can be found where a continuous transition between spin $\frac{1}{2}$ and spin 1 behaviour may be seen depending on a continuous variation of the relative amplitudes of excitation of transitions between 2 or 3 levels.

Even for the above type of experiment we could consider a spin direction which is rotated while the relative population of two levels is changed. To apply our spin-measurement scheme we consider, again on a gedanken-experiment level, a further Stern-Gerlach magnet which separates spatially all three incoming states. Later while one of these states is left unaffected we may recombine the other two states and subject them to the same measurement procedure as above. Thus we may introduce transitions between these two states and we may use one of our modified Stern-Gerlach arrangements to determine the spatial direction for which this two-state subsystem is in an eigenstate. After recombination with the unaffected third state the interference effects may still be observed. Thus, here too it is possible to assign a meaning to the notion of a spatial direction, the rotation of which is associated with the 4π periodicity of the phase factor.

The approach presented here is to some extent complementary to Renninger's considerations of a measurement without disturbance of the measured system²⁰, a concept which has been extended to the spin case by Yoshihuku²¹. The difference is that in Renninger's case the wave packet is actually reduced by the measurement while we consider a measurement without wave packet reduction. Note that there exists a non-vanishing probability that in our modified Stern-Gerlach experiment we may not observe particles in the 'down' detector even if the spin is not in the 'up' eigenstate of the Stern-Gerlach. Evidently that probability can be reduced by increasing the number of particles passing through the apparatus and by carefully orienting the direction of the Stern-Gerlach. Therefore the spin direction can only be determined within a certain error. Our gedanken-experiment shares this inherent statistical nature with other quantum measurements and, to some extent, with classical measurements. Nevertheless, in an interferometer experiment this is a minor point as our modified Stern-Gerlachs actually produce the beams in their respective 'up' eigenstates: the 'down' states are absorbed by the detectors.

The action of our modified Stern-Gerlach magnets including the absorbing detector in one beam path can be described by the projection operator

$$P_{Op} = \frac{1}{2}(I + \mathbf{n}\sigma) \quad (1)$$

Here, $\mathbf{n}$ is a unit vector pointing in the analysing direction of the Stern-Gerlach and σ is the Pauli spin matrix vector. A state is in the 'up' eigenstate of the Stern-Gerlach if

$$\psi = P_{Op}\psi \quad (2)$$

If we write ψ as the spinor

$$\psi = e^{i\epsilon} \begin{pmatrix} a \\ b e^{i\phi} \end{pmatrix} \quad (3)$$

Equation (2) implies that

$$\mathbf{n} = (2ab \cos \phi, 2ab \sin \phi, a^2 - b^2) \quad (4)$$

that is $\mathbf{n}$ has to coincide with the polarization direction of the particles described by ψ .

Disregarding the details of the physics of the beam splitters we can use phenomenologically for the beams inside the interferometer the four-component spinor

$$\psi_{IF} = \begin{pmatrix} \psi_1^+ \\ \psi_1^- \\ \psi_2^+ \\ \psi_2^- \end{pmatrix} = e^{i\epsilon} \begin{pmatrix} a_1 \\ b_1 e^{i\phi_1} \\ a_2 e^{i\chi} \\ b_2 e^{i(\chi+\phi_2)} \end{pmatrix}$$

where, say, ψ_1^+ is the amplitude in the $+z$ -direction in beam 1 and so on. The normalization condition is $a_1^2 + b_1^2 + a_2^2 + b_2^2 = 1$. This four-component spinor describes the states inside the

interferometer, the beams leaving the interferometer are superpositions of its various amplitudes.

In analogy with the above case, the projection operator describing the case where two modified Stern–Gerlachs are in the interferometer, one in each beam path, may be given as

$$P_{IF} = \frac{1}{2} \begin{pmatrix} I + \mathbf{n}_1 \cdot \boldsymbol{\sigma} & 0 \\ 0 & I + \mathbf{n}_2 \cdot \boldsymbol{\sigma} \end{pmatrix} \quad (6)$$

The condition that this operator leaves the wave function invariant

$$\psi_{IF} = P_{IF} \psi_{IF} \quad (7)$$

implies again that

Received 17 June; accepted 24 September 1981.

1. Aharonov, Y. & Susskind, L. *Phys. Rev.* **158**, 1237–1238 (1967).
2. Bernstein, H. J. *Phys. Rev. Lett.* **18**, 1102–1103 (1967).
3. Rauch, H. *et al. Phys. Lett.* **54A** (1975) 425–427.
4. Rauch, H., Wilfing, A., Bauspiess, W. & Bonse, U. *Z. Phys.* **B29**, 281–284 (1978).
5. Rauch, H. & Zeilinger, A. *Hadr. J.* **4**, 1280–1294 (1981).
6. Werner, S. A., Colella, R., Overhauser, A. W. & Eagen, C. F. *Phys. Rev. Lett.* **35**, 1053–1056 (1975).
7. Klein, A. G. & Opat, G. I. *Phys. Rev. Lett.* **37**, 238–240 (1976).
8. Moore, G. T. *Am. J. Phys.* **38**, 1177–1180 (1970).
9. Byrne, J. *Nature* **275**, 188–191 (1978).

$$\mathbf{n}_i = \frac{1}{a_i^2 + b_i^2} (2a_i b_i \cos \phi_i, 2a_i b_i \sin \phi_i, a_i^2 - b_i^2) \quad (8)$$

Or, equivalently, the analysing directions of the individual Stern–Gerlachs have to be oriented parallel to the respective beam polarization directions. The particle state is then left invariant, in particular the relative phase χ in equation (5) between the two interferometer beams is unchanged. Therefore the interference pattern may still be observed.

I thank Professors H. Bernstein, J. Byrne, M. A. Horne, H. Rauch and A. Shimony for critical and useful discussions and Drs A. G. Klein and O. Schärpf for critical reading of the manuscript. Financial support was by Fonds zur Förderung der wissenschaftlichen Forschung (Austria), project no. 4230.

10. Bernstein, H. J. in *Neutron Interferometry* (eds Bonse, U. & Rauch, H.) 231–240 (Oxford University Press, 1979).
11. Mezei, F. in *Neutron Interferometry* (eds Bonse, U. & Rauch, H.) 265–272 (Oxford University Press, 1979).
12. Bernstein, H. J. & Zeilinger, A. *Phys. Lett.* **75A**, 169–172 (1980).
13. Bohm, D. *Quantum Theory*, 593 (Prentice-Hall, New York, 1951).
14. Zeilinger, A. & Shull, C. G. *Phys. Rev.* **B19**, 3957–3962 (1979).
15. Just, W., Schneider, C. S., Ciszewski, P. & Shull, C. G. *Phys. Rev.* **B7**, 4142–4145 (1973).
16. Klempt, E. *Phys. Rev.* **D13**, 3125–3129 (1976).
17. Stoll, M. E., Vega, A. J. & Vaughan, R. W. *Phys. Rev.* **A16**, 1521–1524 (1977).
18. Stoll, M. E., Wolff, E. K. & Mehring, M. *Phys. Rev.* **A17**, 1561–1567 (1978).
19. Kaiser, R. *Can. J. Phys.* **56**, 1321–1332 (1978).
20. Renninger, M. *Z. Phys.* **158**, 417–421 (1960).
21. Yoshihuku, Y. *Mem. Chubu Inst. Technol.* **13-A**, 173–175 (1977).

Detection of monsoon inversion by TIROS-N satellite

M. S. Narayanan & B. M. Rao

Meteorology Division, Space Applications Centre, Ahmedabad-380053, India

Colon¹ and Ramage² have investigated the thermal stratification of the summer monsoon air and presented evidence of a well-defined temperature inversion in the lower atmosphere over the Arabian Sea. This inversion is low (base between 900 and 800 mbar) and strong over the western Arabian Sea and weakens and rises (base at ~700 mbar) towards the coast of India and is not observed east of 70°E, especially during the active monsoon³. The presence of dry warm continental air from Africa and Arabia above the maritime air is thought to be associated with this inversion. This inversion is very important to the rain producing potential of the monsoon current because once the inversion is destroyed there is a favourable stratification for rapid release of moisture upwards leading to precipitation. Observations of the western Arabian Sea inversion features have previously been reported only from *in situ* ship radiosonde and aircraft dropsonde measurements. Although the basic accuracy and the vertical resolution of the present-day satellite sensors cannot delineate the small-scale variations of temperature⁴ such as monsoon inversions we have detected these features from just the TIROS-N derived sea-surface temperatures and the 1,000–850 mbar layer-mean temperatures using a simple differencing procedure. From these temperatures and simultaneous satellite-derived mid-tropospheric water vapour content, we show here the close link between the extent of inversion regions and the convective processes with the Indian monsoon at its different phases.

The present data pertain to the Monex-period (1 May–31 July 1979) TIROS-N satellite results of 14.00 h LT supplied by NOAA Environmental Satellite Service of the USA. These results (about one set in a 2.5° × 2.5° lat.–long. box) include data on the: sea surface temperature (SST); 15 layer-mean atmospheric temperature profile from 1,000 to 0.4 mbar; and 3 level total water vapour content.

The inversions are characterized by the altitude of its base, height extent and temperature departure. Well-marked (height extent >30 mbar and temperature departure >3°C) monsoon inversions in the western Arabian Sea exhibit in the aircraft

dropsonde profiles a negative lapse rate in the altitude regions between 900 and 800 mbar region, above and below which the lapse rates are similar to those of the normal profiles.

Our investigation did not reveal any inversion features from the individual temperature profiles of the satellite data set. However, we now use only the SST and the lowest layer-mean (1,000–850 mbar) temperature, referred to as T_1 and T_2 respectively.

For a temperature profile with an inversion structure at lower levels (below 850 mbar), the layer-mean temperature of the 1,000–850 mbar layer will be larger than that for normal profiles (those which follow a standard lapse rate with height). Assuming further that the sensible heat exchange between the sea and the air above is small and fairly uniform⁵ (implying small air–sea temperature difference), it is to be expected that a horizontal map of the difference, ΔT , of the satellite-derived SST (T_1) and the 1,000–850 mbar layer-mean temperature (T_2), with appropriate time and spatial averaging could reveal the inversion regions. Lower values of ΔT can be interpreted as being associated with regions of inversion. From *in situ* aircraft dropsonde measurements, the values of ΔT for non-inversion profiles in the active monsoon areas range between 4 and 6°C and in inversions these are as low as 0°C in some cases.

Examination of the horizontal variation of the ΔT values thus obtained from satellite data also minimizes the errors that may be present in the original temperatures T_1 and T_2 , whose retrieval involves assumptions of the atmospheric models in the radiative transfer computations.

The aircraft dropsonde measurements during Monex 1979 provided data for *in situ* ground truth comparisons in the Arabian Sea between long. 75°E and 55°E. However, these measurements were available up to 27 June 1979 covering only the onset and active periods of the monsoon. No *in situ* data

Table 1 Comparison of aircraft profiles with satellite data

	Aircraft profiles	Near simultaneous satellite data	
		$\Delta T \leq 2^\circ\text{C}$	$\Delta T \geq 3^\circ\text{C}$
No. of profiles with well-marked inversion below 850 mbar	30	23	7 (for four of them $\Delta T = 3^\circ\text{C}$)
No. of profiles without well-marked inversion	129	0	129

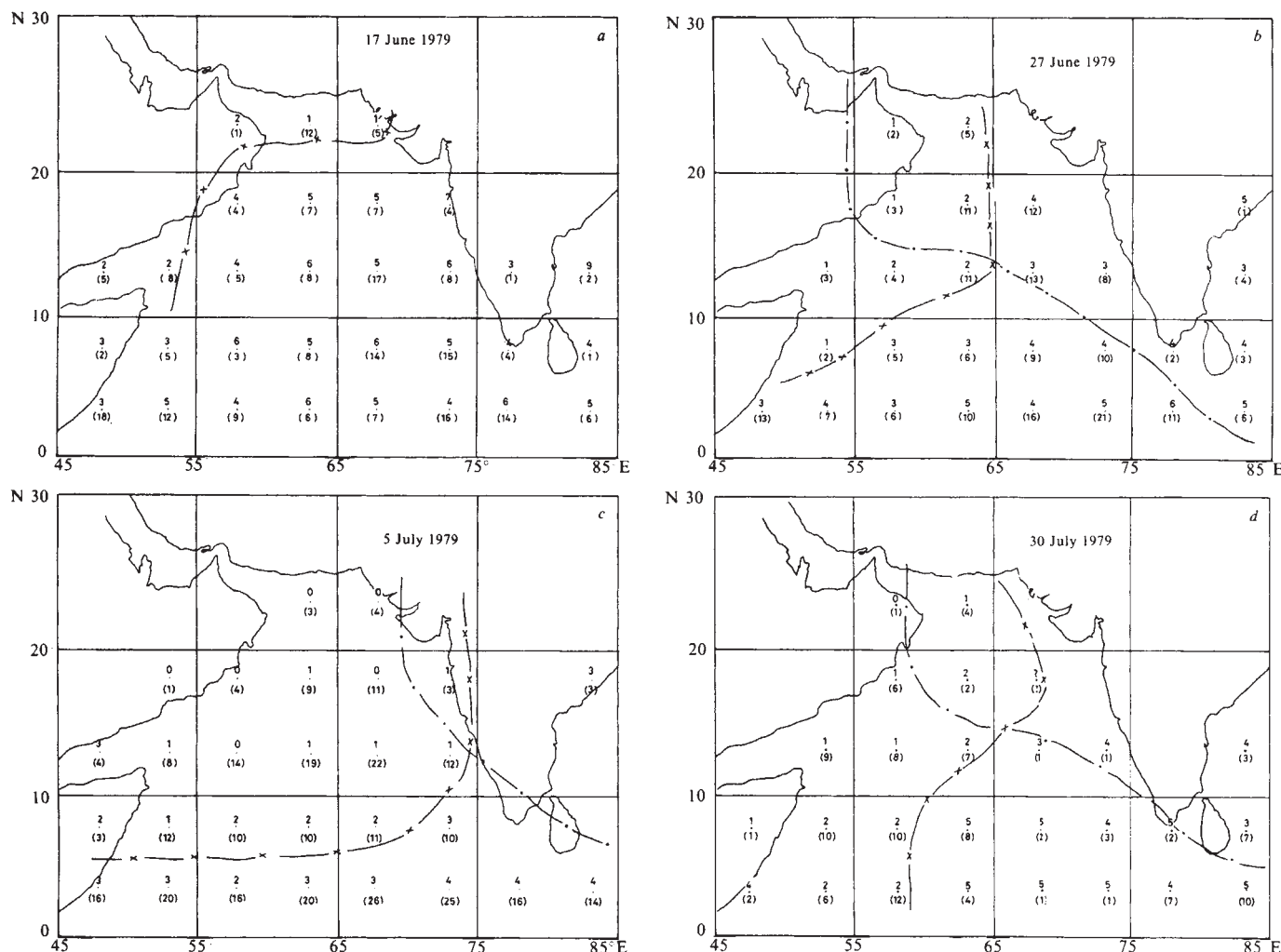


Fig. 1 *a*, ΔT map of the onset phase (3-day average). Values in parentheses denote the number of observations available in the $5^\circ \times 5^\circ$ box for 3 days. —X—, The boundary of the inversion region. *b*, As *a* for the active phase (3-day average). —, The boundary, to the east of which water vapour content in the 700–500 mbar layer is 10 mm or more and to the west less than 10 mm. *c*, As *b* for the weak phase (5-day average). *d*, As *b* for the revival phase (3-day average).

were available for break and revival periods. The *in situ* information available has been used to define inversion regions (up to 850 mbar) in the Arabian Sea.

A total of 85 and 74 aircraft profiles during the onset and active phases respectively were used for comparisons with the satellite data. A fairly large number (20) of well-marked inversions were available only from the profiles of 27 June. The comparative study indicates that inversion-free (up to 850 mbar) regions, as evidenced from dropsonde data, were associated with satellite ΔT values of 3°C or more, and well-marked inversion profiles were associated with ΔT values of 2°C or less.

Table 1 shows that with a high reliability the ΔT values can be used to determine the existence or otherwise of the inversion regions. These comparisons (and the results) indicate that the ΔT variations inferred from our method could be considered accurate to 1°C .

We also attempted to look for inversions above 850 mbar by employing the ΔT maps derived from the 1,000–850 and the 850–700 mbar layer-mean temperature values. No significant horizontal variations of ΔT were observed, indicating that these are either weak, less in height extent or non-existent. The detection of inversions in these regions would need much refined satellite measurements.

The *in situ* statistics of the aircraft dropsonde profiles were used to interpret all available results and ΔT values of 2°C or less have been delineated as regions associated with inversions.

The data at any particular location for any individual day were insufficient for our analysis, mainly due to a large underlap in the tropical regions of about $8\text{--}10^\circ$ in latitude between successive satellite passes, the swath itself being $\sim 10^\circ$ lat., the other major limitation was the cloud cover which inhibited proper temperature retrievals. To help overcome these limitations we adopted a spatial (a $5^\circ \times 5^\circ$ lat.-long. grid box) and time-averaging procedure to investigate the inversion regions in the temperature fields. These optimal nesting and averaging schemes were arrived at from a preliminary examination of the day-to-day plot of the satellite temperature maps and a consideration of the large time and spatial scales associated with the monsoon inversions.

Figure 1 presents the ΔT maps during the different phases of the monsoon of 1979. Values of ΔT indicated over land correspond to the $5^\circ \times 5^\circ$ box central location. The contribution is due entirely to those soundings from over the sea. A feature which persists throughout the observational period is the tendency of ΔT values to increase slowly from west to east as the monsoon becomes established.

The map of the pre-onset time (not reproduced here) shows ΔT values at all places more than 3°C indicating no inversion features. Figure 1*a* shows the inversion being set up in the western Arabian Sea at the onset time of the monsoon. The low ΔT values start covering a larger part of the Arabian Sea (but confined still to west of 65°E) at the most active phase of the monsoon (Fig. 1*b*). These results agree with those reported by

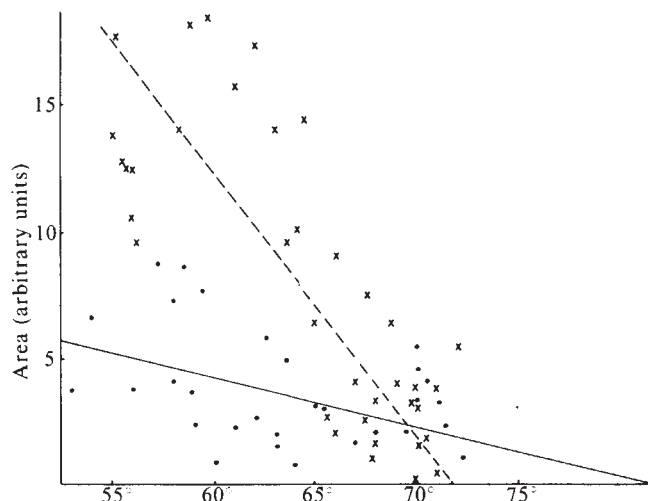


Fig. 2 Strength of inversion at onset and active phases from *in situ* data. —, Onset phase (●); ---, active phase (×).

Sen and Das⁶ and Ghosh *et al.*³, who have shown from limited Monex (1979) and Ismex (1973) *in situ* data respectively, that the inversion regions are confined to west of 65° E during the active periods.

The area included between the normal profile (with standard lapse rate) and the observed profile (with or without inversion) is a good *in situ* measure of the inversion strength, as it incorporates both the height extent and temperature departure features of inversion. Figure 2 shows a scatter plot of the area so calculated between surface and 700 mbar for the *in situ* dropsonde profiles for the onset and active periods. The gradual increase of area from east to west and a sharp increase of the slope from the onset to the active phase as revealed by *in situ* observations of this plot agree well with similar features revealed by the satellite ΔT maps (Fig. 1a, b).

We also report here observations of the inversion features available from the ΔT maps during the weak and the revival phases of the monsoonal activity. No observations *in situ* or otherwise have previously been reported of the inversion features during these periods of the monsoon. Our analysis shows that during times of weak monsoon activity the inversion regions intrude east of 65° E (Fig. 1c) and even encroach the west coast of India. The revival of monsoon after 29 July is revealed by the ΔT map (Fig. 1d), with a shifting to the west of ΔT values, and these being 3 °C or more east of 65° E.

Figure 1b–d also shows the 10-mm total water vapour content contour corresponding to the 700–500 mbar level drawn from the data of the TIRDS-N observations. East of this contour the values are higher and to the west lower than 10 mm. Figure 1b–d reveals a clear oscillation from west to east and then back to west longitudes of this contour between the active, weak and revival phases of the monsoon.

We have not observed such clear oscillations of the lower level (1,000–700 mbar) water vapour content contours. The observation of the middle level moisture content and its change during the different phases of the monsoon (illustrating the convective activity) are consistent with our observations of the east–west shifts of the low level inversions, and show the close link of these two phenomena with the general monsoon circulation.

We thank Dr T. A. Hariharan, Head, Meteorology Division, for encouragement and guidance; also Professor J. N. Desai, Drs P. S. Sehra and P. S. Desai for useful discussions.

Received 6 July; accepted 14 October 1981.

1. Colon, J. A. *Ind. J. met. Geophys.* **15**, 183–200 (1964).
2. Ramage, C. S. *J. Atmos. Sci.* **23**, 144–150 (1966).
3. Ghosh, S. K., Part, M. C. & Dewan, B. N. *Tellus* **30**, 117–125 (1978).
4. Philips, N., McMillin, L., Gruber, A. & Wark, D. *Bull. Am. met. Soc.* **60**, 1188–1197 (1979).
5. Cayan, D. R. *Mn. Weath. Rev.* **108**, 1293–1301 (1980).
6. Sen, P. & Das, I. P. *FGGE Rep. Ser.* **9A**, 225–233 (1980).

Steady-state magma discharge at Etna 1971–81

G. Wadge

Lunar and Planetary Institute, 3303 NASA Road One, Houston, Texas 77058, USA

J. E. Guest

University of London Observatory, Mill Hill Park, London NW7 2QS, UK

Throughout the past decade Mount Etna has been in almost continuous activity and even during periods of repose incandescent lava has often been visible in at least one of the summit vents. Using observations by Italian, British and French volcanological teams we have estimated the volumes of lava produced by each eruption from 1971 to July 1981. The computed output of magma for this period approximates to a rate of $0.7 \text{ m}^3 \text{ s}^{-1}$. We now compare this with our output rate estimates for Etna's historic past. The steady-state nature of the output during the past decade has implications for the interpretation of the volcano's internal plumbing and the petrology of its lavas, and the assumption that this state will be maintained allows us to discuss the timing and magnitude of future eruptions.

In 1975 we presented evidence¹ for the rate at which magma appears at the summit and at the flank vents of Etna during the period 1966–74 and for the rate at which it erupted solely on the flanks of the volcano over a longer period (1535–1974). We now extend this record of volumetric discharge of magma to July 1981 (Table 1).

During the period 1971–July 1981 there were six eruptions involving flank vents (1971, 1974, August 1978, November 1978, 1979 and 1981). The rest of the activity was at the summit vents (mainly north-east and south-east craters) and at secondary vents lower down the volcano (mainly to the north), some of which were fed by lava tubes from the summit vents. Many of the eruptions were short, explosive, high effusion rate events such as those in the latter half of 1977. Other eruptive episodes involved continuous effusion for months at a time² such as those typifying the period September 1974 to June 1976. Our volume data (Table 1) come from various sources and are not corrected for vesicularity. The most accurate of these data are the result of theodolite surveying of the summit region^{3–5}. Most lava flow volumes lower on the flanks have been estimated from their areas and average thicknesses. The accuracy of these less precise estimates is probably better than $\pm 50\%$.

The cumulative volume plot for January 1971–July 1981 (Fig. 1) has a linear trend which gives an output rate⁶ of $0.7 \text{ m}^3 \text{ s}^{-1}$ ($0.67 \text{ m}^3 \text{ s}^{-1}$). For the period September 1974–June 1976 when effusion was almost continuous, though variable over short periods of time, the rate was indistinguishable ($0.67 \text{ m}^3 \text{ s}^{-1}$) from the overall 1971–July 1981 rate. After a period of repose during 1972–73 following the 1971 flank eruption the cumulative volume curve has kept within the limits shown in Fig. 1, which represents a range of volume of about $25 \times 10^6 \text{ m}^3$. There is a tendency for a change in style of eruption during the second half of this steady-state period (July 1977–July 1981), coupled with a slightly increased gradient of the curve, but this appears to be of secondary importance.

We interpret these features of Fig. 1 as follows. From 1971 to July 1981 magma has been entering and leaving Etna at a rate of $0.7 \text{ m}^3 \text{ s}^{-1}$. At times this steady state has been achieved exactly (within the precision of our data) as in 1975 and is equivalent to the behaviour of Kilauea in 1952, 1967–68 and 1969–71 (see ref. 7). At other times magma is either stored in the volcano before release or the magma arrives in discrete batches which

Table 1 Volumes of the eruptions of Etna 1971–July 1981

Eruption	Volume ($\text{m}^3 \times 10^6$)	Σ Volume ($\text{m}^3 \times 10^6$)	Ref.	Eruption	Volume ($\text{m}^3 \times 10^6$)	Σ Volume ($\text{m}^3 \times 10^6$)	Ref.
1971 5 April–12 June	50	50	1	1977 10–13 December	0.5	129	
1971 June–September	7	57	1	1977 18 December	1	130	
1973 October	10	67	1	1977 24–25 December	1	131	
1974 31 January–17 March	6	73	20	1977 29 December	1	132	4, 24–27
1974 29 September–24 February	7	80	3	1978 2–4 January	2.5	134.5	
1975 24 February–12 September	13	93	2, 21, 22	1978 5 January	0.5	135	
1975 12 September–28 November	5	98	3, 23	1978 7 January	0.5	135.5	
1975 29 November–17 June	12	110	21, 23	1978 25–28 March	3	138.5	28
1976 20 August–8 January	6	116	21, 24, 25	1978 29 April–5 June	23	161.5	4, 24
1977 16–22 July	1	117		1978 23–30 August	3	164.5	4, 24, 29
1977 5–6 August	1	118		1978 23–30 November	5	169.5	4, 24, 29
1977 14 August	4	122		1979 3–9 August	12	181.5	24, 30
1977 24 August	0.5	122.5		1980 April	4	185.5	31, 32
1977 2–4 November	1	123.5	4, 24–27	1980 1 September	7	192.5	33
1977 7–8 November	1	124.5		1980 6 September	3	195.5	33, 34
1977 22 November	1	125.5		1981 5–7 February	4	199.5	35
1977 25 November	1	126.5		1981 17–20 March	20	219.5	36
1977 6 December	2	128.5					

All volumes estimated to the nearest $1 \times 10^6 \text{ m}^3$ except eruptions from 16 July 1977 to 7 January 1978 which are estimated to the nearest $0.5 \times 10^6 \text{ m}^3$.

results in a punctuated pattern of magma release. The maximum capacity of either of these processes (or a combination of them) is $\sim 25 \times 10^6 \text{ m}^3$.

Recognition of this rate of magma discharge and the capacity of the episodic process(es) since 1974 (Fig. 1) allows us to suggest how Etna will behave in the immediate future. If the cumulative volume curve is to remain within the empirically defined limits Etna must erupt again before about May 1982. At that time up to $25 \times 10^6 \text{ m}^3$ of magma could be expected, but an eruption volume of less than about $17 \times 10^6 \text{ m}^3$ could be expected from an eruption in January 1982 (Fig. 1). These are not predictions of specific events but general forecasts of the timing and magnitude of future eruptions based on the behaviour of

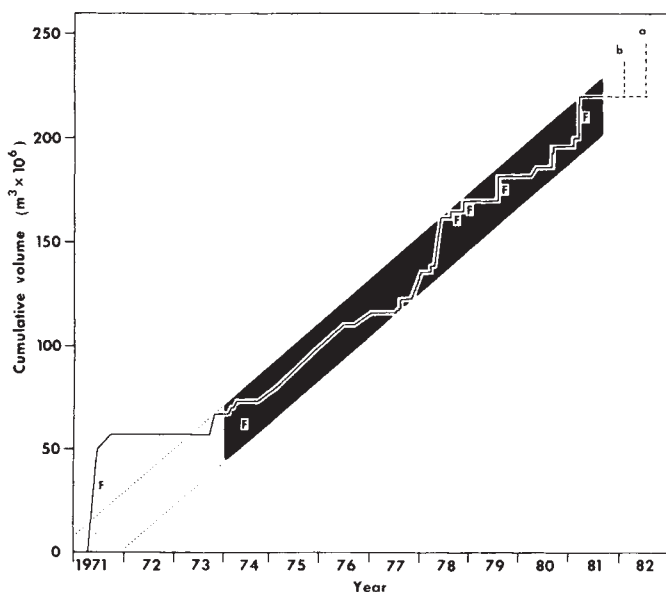


Fig. 1 Plot of the cumulative volume of magma erupted at Etna from 1971 to July 1981 listed in Table 1. F indicates eruptions involving flank vents. From January 1974 to July 1981 the cumulative volume curve has remained within the limits shown by the shaded area which corresponds to the average discharge rate calculated for the entire period 1971–July 1981 ($0.7 \text{ m}^3 \text{ s}^{-1}$). If the volcano continues to behave in this way an eruption is expected to occur before approximately May 1982 ± 3 months. An eruption at that time should not exceed a volume of about $25 \pm 6 \times 10^6 \text{ m}^3$ (a). The maximum volume of an eruption in, say, January 1982 would be about $17 \times 10^6 \text{ m}^3$ (b). If instead there is a larger volume eruption (as in 1971) a correspondingly long period of repose is expected to follow.

the volcano during the past seven years. The accuracy of such forecasts depends on the error involved in the volume estimates used in constructing the curve. Random errors in individual estimates (probably a few 10%) will tend to cancel in the cumulative volume curve and reduce the uncertainty in the limits of the shaded envelope of Fig. 1. Unknown systematic errors will affect the gradient of the curve. The above forecast depends in particular on the accuracy of the last eruption volume of the curve (March 1981), which is about $\pm 10\%$ ($20 \pm 2 \times 10^6 \text{ m}^3$). Combined with an assumed uncertainty of the steady-state limits (shaded envelope) of $\pm 20\%$ this produces an uncertainty of about ± 3 months and about $\pm 6 \times 10^6 \text{ m}^3$ in the maximum repose period and volume of the next eruption.

The 1971 eruption does not fit the above pattern. From January 1966 until April 1971 summit effusion was very similar in style and magnitude to the 1975 activity^{8,9} and the rate of $0.7 \text{ m}^3 \text{ s}^{-1}$ is also probably applicable to the 1966–71 period. If this is true then the 1971 eruption represents an event which involved an unusually large volume of magma and after which there was an unusually long repose period. The proposed steady-state rate of discharge from 1966 to 1981 of $0.7 \text{ m}^3 \text{ s}^{-1}$ is equated with the rate of supply of magma from the lower crust and the 1971 eruption volume is part of that supply rather than originating in some additional reservoir or source region. The large volume implies that either a high-level reservoir was drained or that an unusually large batch of magma rose into the volcano, but that both were part of a steady supply system which required a subsequent period of recharge during 1972–73. If our assumption about the steady-state output rate from 1966 to April 1971 is incorrect then this weakens, but does not invalidate, the case for the 1971 eruption magma being part of the steady-state supply.

These conclusions reinforce the previously noted¹ discrepancy between the modern discharge rate and the rate calculated from the flank eruptions of the past 200 yr ($0.17 \text{ m}^3 \text{ s}^{-1}$). We suggested¹ that summit eruptions, whose volumes cannot be estimated for this period, might increase this latter value to $0.26 \text{ m}^3 \text{ s}^{-1}$. However, the modern (1971–81) ratio of summit to flank magma volume of 56:44, if applied to the record of the past 200 yr, gives a combined summit and flank output rate of $0.39 \text{ m}^3 \text{ s}^{-1}$. There has been one historic period (1610–69) during which magma erupted at a rate ($0.83 \text{ m}^3 \text{ s}^{-1}$) comparable with the 1971–81 rate¹. Between 1610 and 1669 there was no recorded summit effusion and the lavas of the very long-lasting flank eruptions were characterized by particularly large plagioclase phenocrysts¹⁰. The large phenocrysts may imply longer residence times for magma at high levels than at present. Normal access to the summit vents may have been

blocked at this time and all the magma from depth may have found its way to the surface through flank vents.

The steady-state discharge of magma at Etna provides a framework within which some of the findings of recent research on the volcano can be correlated. The major element geochemistry of historical Etna lavas has shown little variation with time^{10,11}. A steady-state discharge rate must be an important factor in determining the chemical homogeneity of these hawaiite lavas. Small variations in residence time at high levels in the feeding fissures are probably capable of producing the observed limited variations in chemistry. Direct monitoring of SO₂ by correlation spectrometry in the plume of Etna during June 1975¹², August 1977¹³ and July 1978¹⁴ has revealed a characteristic discharge rate for this species varying between 1,000 and 5,000 tonnes per day depending on the level of activity. A value within this range may represent the steady rate of release of SO₂ from a hawaiite magma flux of 0.7 m³ s⁻¹, which is ~0.6–3 wt.% of the mass of erupted magma (assuming the specific gravity of the magma is 2,650 kg m⁻³). However, this range is much higher than the sulphur contents of most sulphide-saturated basaltic liquids (<0.25 wt% S)¹⁵. If the spectrometer values are accurate then the Etna magmas are either exceptionally rich in sulphur or a larger volume of magma may be degassing than appears at the surface at the steady-state rate.

Various techniques for monitoring the surface deformation of Etna have been since 1971, including precise levelling, electronic distance measurements and several types of tiltmeters^{16–19}. The most substantial deformations were measured from 1971 to 1974 during the period of recovery from the 1971 eruption¹⁶. Subsequent measured deformation has been relatively minor¹⁰. This suggests that there is no elastic reservoir system within the volcano that is sensitive to the discharge of up to 25 × 10⁶ m³ of magma within the steady-state limits of Fig. 1.

We thank the many scientists whose observations on Etna over the past decade made this work possible. J.E.G. acknowledges financial support from NERC. G.W. is a visiting scientist at the Lunar and Planetary Institute which is operated by the Universities Space Research Association under contract NASW-3389 with NASA.

Received 18 August; accepted 6 October 1981.

- Wadge, G., Walker, G. P. L. & Guest, J. E. *Nature* **255**, 385–387 (1975).
- Pinkerton, H. & Sparks, R. S. J. *J. Volcan. Geotherm. Res.* **1**, 167–182 (1976).
- Murray, J. B. *U.K. Research on Mount Etna 1975–76*, 9–12 (Royal Society, London, 1977).
- Murray, J. B. *U.K. Research on Mount Etna 1977–79*, 33–42 (Royal Society, London, 1980).
- Murray, J. B. *U.K. Research on Mount Etna 1974*, 46 (Royal Society, London, 1975).
- Wadge, G. *Nature* **288**, 253–255 (1980).
- Swanson, D. A. *Science* **175**, 169–170 (1972).
- Walker, G. P. L. *Nature* **213**, 484–485 (1967).
- Guest, J. E. *Phil. Trans. R. Soc. A* **274**, 64–78 (1973).
- Guest, J. E. & Duncan, A. M. *Nature* **290**, 584–586 (1981).
- Cristofolini, R. *Phil. Trans. R. Soc. A* **274**, 17–35 (1973).
- Haulet, R., Zettwoog, P. & Sabroux, J. C. *Nature* **268**, 715–717 (1977).
- Malinconico, L. L. Jr *Nature* **278**, 43–45 (1979).
- Faivre-Pierret, R., Martin, D. & Sabroux, J. C. *Bull. Volcan.* **43**, 473–486 (1980).
- Anderson, A. T. *Bull. geol. Soc. Am.* **85**, 1485–1492 (1974).
- Wadge, G. *J. Volcan. Geotherm. Res.* **1**, 237–264 (1976).
- Murray, J. B., Guest, J. E. & Butterworth, P. S. *Nature* **266**, 338–340 (1977).
- Murray, J. B. & Guest, J. E. *U.K. Research on Mount Etna 1977–79*, 18–20 (Royal Society, London, 1980).
- Henbest, S. N., Mills, A. A. & Ottey, P. J. *Volcan. geotherm. Res.* **4**, 133–149 (1978).
- Bottari, A., Lo Giudice, E., Patane, G., Romano, R. & Sturiale, C. *Riv. Miner. Sicil.* **154–6**, 175–198 (1975).
- Guest, J. E. & Murray, J. B. *U.K. Research on Mount Etna 1975–76*, 56–57 (Royal Society, London, 1977).
- Booth, B., Sparks, R. S. J. & Walker, G. P. L. *U.K. Research on Mount Etna 1974*, 69–71 (Royal Society, London, 1975).
- Romano, R. & Sturiale, C. *Bull. Volcan. Erupt.* **16**, 9–10 (1978).
- Guest, J. E. & Murray, J. B. *U.K. Research on Mount Etna 1977–79*, 50–52 (Royal Society, London, 1980).
- Romano, R. *Bull. Volcan. Erupt.* **17**, 8–9 (1980).
- SEAN Bull.* 2/11, 1 (Smithsonian Institution, 1977).
- SEAN Bull.* 2/12, 3 (Smithsonian Institution, 1977).
- SEAN Bull.* 3/3, 2 (Smithsonian Institution, 1978).
- Mackey, M. & Scott, S. C. *U.K. Research on Mount Etna 1977–79*, 43–44 (Royal Society, London, 1980).
- SEAN Bull.* 4/8, 12–13 (Smithsonian Institution, 1979).
- SEAN Bull.* 5/7, 8–9 (Smithsonian Institution, 1980).
- Kasser, M., Delorme, H., Kieffer, G. & Tanguy, J.-C. *C.R. hebd. Acad. Séanc. Sci., Paris* **287**, 1191–1194 (1978).
- SEAN Bull.* 5/8, 6–7 (Smithsonian Institution, 1980).
- SEAN Bull.* 5/9, 5 (Smithsonian Institution, 1980).
- SEAN Bull.* 6/2, 6–7 (Smithsonian Institution, 1981).
- SEAN Bull.* 6/3, 3–5 (Smithsonian Institution, 1981).

Charnockite and granite formation and influx of CO₂ at Kabbaldurga

C. R. L. Friend

Department of Geology and Physical Sciences, Oxford Polytechnic, Oxford OX3 0BP, UK

The discovery, at Kabbaldurga quarry, Karnataka, south India^{1–3}, of patches of charnockite apparently in an arrested state of development has refocused attention on the mechanism of charnockite formation. In particular, the relative roles of CO₂ and H₂O during high grade metamorphism and charnockite development can be studied in such an area. During an influx of a CO₂-rich volatile phase, H₂O may be liberated as a result of the reduction in the state of hydration by the breakdown of hydrous minerals⁴. The *P–T* conditions in which this occurs are important because, with an increase in XH₂O, conditions which may allow anatexis could be initiated in advance of such a CO₂-rich phase. Indeed, Weaver⁵ has suggested that the K-rich acid charnockites found at Pallavaram, Madras⁶, developed due to changes in volatile composition before the onset of charnockite conditions. The exposure at Kabbaldurga quarry is interpreted here as displaying evidence that such a process has taken place.

During recent field work in the Archaean (~2,700 Myr) gneiss terrain between Mysore and Bangalore, Karnataka, south India, the southern end of the Closepet Granite⁷ including the area of Kabbaldurga was examined. This granite is a substantial body with a linear extension northwards for some 200 km to just south of Bellary. The country rocks into which the granite has been emplaced are dominantly polyphase, grey Archaean gneisses—the Peninsular Gneiss complex. In this area the gneissic complex has been subjected to upper amphibolite and granulite facies conditions of metamorphism. The age of the Peninsular Gneiss complex precursors is not well known, but is at least 2,600 Myr (ref. 8), an age which reflects a metamorphic event. More recently, an Rb–Sr whole rock isochron age of 3,358 Myr has been obtained for part of the complex to the west of Bangalore⁹. Two homogeneous foliated granitoid masses, which also form a portion of the Peninsular Gneiss, at Chickmagalur and at Chitradurga, have given Rb–Sr ages around 3,080 Myr and 2,790 Myr respectively (S. Moorbath and P. N. Taylor, personal communication). These ages for components of the gneiss complex are consistent with that of 2,670 Myr (ref. 10), interpreted as a metamorphic age, which has been obtained from the charnockite and the gneisses at Kabbaldurga and 2,615 Myr for the general granulite facies event⁸ of the area.

Among the studies of the Closepet Granite^{7,11,12} the observations that most closely approach those made during the present investigation are those of Radhakrishna⁷, who differentiated the granite into four main varieties. Briefly, the present investigation suggests that the granite is itself a polyphase body and comprises several pinkish and/or greyish, coarse-grained, porphyritic and non-porphyritic granites (using the nomenclature of Streckeisen¹³) and various pegmatitic and aplitic phases. In some portions of the granite enclaves of the country rocks are present. In these areas it is evident that pink alkali-feldspar megacrysts present in the porphyritic granites are not phenocrysts but are of replacive origin because the megacrysts are developed in and across the margins of some of the enclaves. Additionally in some areas, white alkali-feldspar megacrysts have been partially replaced by veins and patches of pink alkali-feldspar. This accords with the observations of Radhakrishna⁷ and with the hypothesis¹¹ that metasomatism is involved in the production of the granite. There is frequently a distinct orientation of both types of megacrysts which is parallel to a variably developed foliation picked out by an alignment of biotite and to the northerly trend of the granite body. This feature suggests that the megacrysts have developed in response to late-stage K-rich fluids in a still active stress zone. This fluid could also be the source of the enrichment in Th, Rb and Pb as

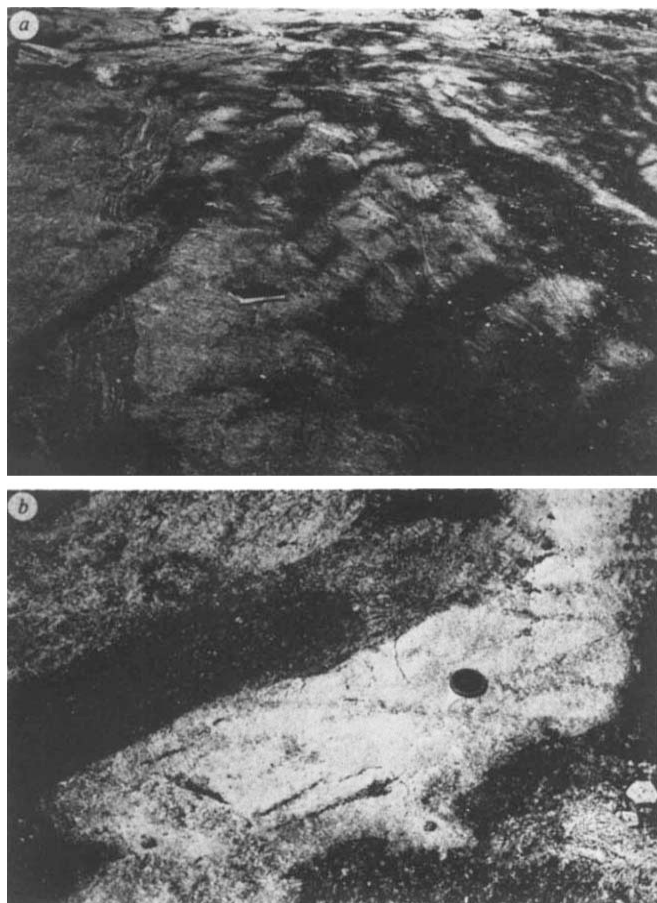


Fig. 1 The penecontemporaneous relationships of granite and charnockite. *a*, The nebulitic development of charnockite after the formation of granite. In the foreground and left-hand side grey migmatitic gneiss is cut by a sheet of granite. The right-hand contact of this sheet is gradational into the gneiss. Both of these lithologies have been overprinted by the nebulitic development of charnockite (see ref. 1) which forms the black areas. Running diagonally across the top right-hand corner of the photograph is an area of charnockite which occurs as anastomosing veins overprinting an area dominantly of granite. *b*, A thin sheet of granite (~28 cm) without charnockite development, cutting and containing two schlieren of charnockite developed from grey gneiss. Above the granite sheet is a nebulitic contact between an area of charnockite and gneiss in which charnockite has begun to develop.

well as K reported from the northern portion of the granite body¹². Additionally, the granite has a broad overall zonation⁷ which is parallel to the northerly trend of the body.

On the southeastern side the rocks comprising the marginal contact zone of the granite are of particular interest. This contact zone consists of a series of intermixed, polyphase granitic sheets and veins (together with pegmatitic and aplitic rocks) enclosing and cutting enclaves of Peninsular Gneiss. In this area the Peninsular Gneiss consists of typical grey, banded gneisses with a few horizons of more basic material and some bodies of essentially unfoliated, more homogeneous gneiss. In the contact zone the enclaves display a range of textural and structural modification from relatively angular, unmodified gneisses into highly banded rocks passing into nebulitic and inhomogeneous, coarse-grained material which lacks a gneissic fabric and may be differentiated into neosome and palaeosome. Such material exhibits the structures and textures of migmatites^{14,15}. The transitions from these migmatitic rocks into more homogeneous granitic rocks are attributed to the process of anatexis and may be described as homogeneous and inhomogeneous diatexites (see ref. 15). From the field evidence in these areas I conclude that the origin of the Closepet Granite is due to partial melting in the Peninsular Gneisses. Crawford⁸ has determined the initial ⁶⁷Sr/⁸⁶Sr ratio of the granite to be 0.705 ± 0.0014 which,

for ~2,400 Myr, suggests a derivation by remobilization of pre-existing sialic material. The importance of this observation becomes apparent in the Kabbaldurga quarries where the relationships of the Closepet Granite to the formation of charnockitic rocks may be observed. The association of charnockitic rocks and migmatites is a feature which has long been known from other regions¹⁶. Kabbaldurga quarry is an area where data regarding this interesting relationship may be collected.

On close investigation the patchy development of charnockite can be seen to have occurred not only in the Peninsular Gneisses as has previously been described¹⁻³ but also in some of the rocks which form a part of the Closepet Granite. From field evidence the development of the charnockitic patches seems to be penecontemporaneous with the development of the granite. Veins and sheets of granite, without charnockitic assemblages, cut charnockitic patches in the gneisses yet are themselves cut by other granitic veins with charnockite development in them (Fig. 1). Note that the areas of gneiss in which charnockite has developed have not undergone partial melting³. However, in some of the grey, biotite gneisses around the patches of charnockite partial melting to produce granitic material has taken place. These rocks contain the structures and textures which closely resemble those found elsewhere along the southeastern marginal zone of the granite and which are interpreted as migmatitic and anatexitic features. The onset of charnockite formation at Kabbaldurga is thought to have taken place at between 3 and 5 kbar at 600–700 °C (ref. 3), (though this is a low estimate compared with the values suggested for the surrounding charnockite formation¹⁷, the mineral assemblages present do not allow a more accurate estimate to be made) and to have been initiated by an influx of a mantle-derived volatile phase rich in CO₂. Such a mechanism has now been more widely applied to granulite facies rocks^{5,18} to explain in particular the formation of hypersthene-bearing assemblages without a change in the existing *P-T* conditions. Whilst Janardhan *et al.*³ argue that a large rise of temperature is not required for the formation of charnockite at Kabbaldurga, the invading volatile phase may have brought a source of heat with it particularly as the volatile phase is considered to be mantle-derived³. Schuiling and Kreulen¹⁹ attribute considerable potential to an invading volatile phase with respect to the release of heat and the initiation of anatexis. Although data enabling the application of their model calculations are not obtainable, any influx of heat will assist the breakdown of the hydrates present and the production of the observed migmatitic features. At Kabbaldurga, therefore, during such an advance of CO₂, the hydrous minerals—in this case biotite and amphibole—have broken down, without partial melting, to give hypersthene-bearing, non-foliated, charnockitic rocks. The H₂O released by this breakdown would have immediately joined the volatile phase which would continue its upwards migration. Therefore, at any point with the continued breakdown of hydrous phases more H₂O would be added to the volatile phase, eventually diluting the composition of the volatiles until such a time as the CO₂ front advanced through that point. Although the invading volatile phase is K-rich³ there is no unequivocal evidence that this is a different volatile phase from that which gives rise to the growth of alkali feldspar megacrysts in the granite. I believe, however, that this latter phenomenon is a late-stage crystallization effect—in common with many other granites—and that the two volatile phases are not directly related.

The exposure at Kabbaldurga seems to have been in a rather special position, at the very front of the CO₂ advance, where the introduced volatile phases are subject to dilution by the released H₂O. The H₂O released by the breakdown of the hydrous phases has had serious consequences. In some areas of the grey gneisses not affected by charnockite formation, the granite solidus was reached and partial melting took place. Then, after the formation of the granite, the CO₂ front advanced into some of the areas where partial melt had accumulated and has formed K-rich charnockite. That there are areas of Closepet Granite in which charnockite has developed demonstrates that the two

events were penecontemporaneous. Weaver⁵, in discussing the generation of the K-rich charnockites at Pallavaram, considered that they represent an unusual component of an otherwise fairly typical Archaean grey gneiss—granulite terrain. He suggested that these K-rich rocks had developed due to a change of the fluid composition which allowed metasomatism and partial fusion to occur before the formation of charnockite by the advance of a CO₂ front. At the present level of erosion, the exposure is interpreted as having been well below the former position of the CO₂ front and all the rocks have been completely converted to charnockite.

At Kabbaldurga it appears that there is evidence not only of the arrested process of charnockite formation, but also of the process of crustal fusion induced by an accumulation of H₂O in advance of an ascending CO₂-rich volatile phase. Thus, this area of the Peninsular Gneiss complex and the Closepet Granite is fundamental to the understanding of the processes of both charnockite formation and granite formation deep in the crust.

A British Council ALIS award is acknowledged enabling the continuation of work begun earlier under a YSES fellowship. I also thank Professor C. Naganna and Dr B. Mahabaleswar, of Bangalore University, for financial assistance and for help in the field respectively and Drs B. P. Radhakrishna and A. S. Janardhan for discussions on the Closepet Granite and the Kabbaldurga area and M. Brown, V. R. McGregor and C. G. Topley for helpful comments.

Received 12 May; accepted 6 October 1981.

- Pichamuthu, C. S. *Nature* **188**, 135 (1960).
- Ramiengar, A. S., Ramakrishnan, M. & Viswanatha, M. N. *J. geol. Soc. Ind.* **19**, 411–419 (1978).
- Janardhan, A. S., Newton, R. C. & Smith, J. V. *Nature* **278**, 511–514 (1980).
- Heier, K. S. *Phil. Trans. R. Soc. A273*, 429–442 (1973).
- Weaver, B. L. *Contr. Miner. Petrol.* **71**, 271–279 (1980).
- Weaver, B. L., Tarney, J., Windley, B. F., Sungavarman, E. B. & Venkata Rao, V. in *Archaean Geochemistry* (eds Windley, B. F. & Naqvi, S. M.) 177–204 (Elsevier, Amsterdam, 1978).
- Radhakrishna, B. P. *Mysore geol. Ass. Bangalore Spec. Bull.* **3**, 1 (1956).
- Crawford, A. R. *J. geol. Soc. Ind.* **10**, 117–166 (1969).
- Beckinsale, R. D., Drury, S. A. & Holt, R. W. *Nature* **283**, 469–470 (1980).
- Venkatasubramanian, V. S. & Narayanaswamy, R. *J. Ind. Inst. Sci.* **56**, 19–42 (1974).
- Divakara Rao, V., Aswathanarayana, U. & Qureshy, M. N. *J. geol. Soc. Ind.* **13**, 1–12 (1972).
- Divakara Rao, V., Aswathanarayana, U. & Qureshy, M. N. *Mineral. Mag.* **38**, 678–686 (1972).
- Streckeisen, A. L. *Earth Sci. Rev.* **12**, 1 (1975).
- Menhart, K. R. *Migmatites and the Origin of Granitic Rocks* (Elsevier, Amsterdam, 1968).
- Brown, M. *Proc. geol. Ass.* **84**, 371–382 (1973).
- Quensel, P. *Ark. Mineral. Geol.* **1**, 227–322 (1951).
- Perkins, D. & Newton, R. C. *Nature* **292**, 145–146 (1981).
- Newton, R. C., Smith, J. V. & Windley, B. F. *Nature* **288**, 45–49 (1980).
- Schilling, R. D. & Kreulen, R. *Earth planet. Sci. Lett.* **43**, 298–302 (1979).

Limitations on the scale of mantle heterogeneities under oceanic ridges

Bernard Dupré, Bernard Lambret,
Dominique Rousseau & Claude J. Allègre

Laboratoire de Géochimie et Cosmochimie, Institut de Physique du Globe et Département des Sciences de la Terre, Universités Paris VI et VII, 4 place Jussieu, 75230 Paris Cédex 5, France

The scale of mantle heterogeneities has been debated ever since the first observations of isotopic variations along mid-oceanic ridges^{1,2}. Subsequent studies on lead and strontium isotopic variations along the Mid-Atlantic Ridge^{3–5} have shown that these isotopic compositions may vary with a major wavelength of ~100–1,000 km. The immediate question is that of the scale (if any?) at which one may consider a piece of suboceanic mantle under a ridge to be homogeneous. We have studied two segments of oceanic ridges at the scale of a few kilometres. One of these portions of oceanic ridge, the CYAMEX zone in East Pacific Rise is typical, with a strongly 'depleted' chemistry. The other one, with an 'intermediate' chemistry, is the FAMOUS zone of the Mid-Atlantic Ridge. Our results show isotopic homogeneity for both zones even though small Pb variations persist on a small scale.

The general geology of the FAMOUS area has been described elsewhere^{6,7}. In the central valley of this slow spreading ridge (2 cm yr⁻¹), we can distinguish two types of structures, real volcanoes such as Mount Venus, and lava flows, probably of fissural type. Several samples were collected from each structure. Figure 1 shows the location of the 16 samples analysed; the area covered was a few square kilometres.

The CYAMEX zone is located on the East Pacific Rise at a latitude close to 21° N. The mean spreading rate of this ridge is ~6 cm yr⁻¹. In contrast with the FAMOUS area, the CYAMEX zone presents no clearly defined central valley. Detailed tectonic⁸, petrographic⁹ and geochemical¹⁰ studies are available. The area is characterized by more abundant fluid lavas than in FAMOUS, some in the form of the peculiar 'lava lakes'⁸. Hydrothermal circulation has been discovered in this area where the massive sulphide or manganese deposits originate¹¹.

Measurements and normalization of lead and strontium isotope ratios were completed using techniques described in refs 12, 13: the results are presented in Table 1.

In the FAMOUS area the strontium isotope values for whole rocks vary from 0.70287 to 0.70347. Such variations may *a priori* be ascribed to several causes. After leaching, using O'Nions and Pankhurst's¹⁴ differential solution technique, samples yield values which are very close to 0.70287, in agreement with analyses by White^{15,16} and O'Nions and Pankhurst¹⁴ for neighbouring areas. Therefore, the variation mentioned above can be attributed to seawater alteration and within the experimental error, the isotopic composition of Sr seems to be very constant.

The lead isotope values are rather uniform and close to mean values of ²⁰⁶Pb/²⁰⁴Pb = 18.7, ²⁰⁷Pb/²⁰⁴Pb = 15.53, ²⁰⁸Pb/²⁰⁴Pb = 38.20 with a maximum variation of 0.2 for ²⁰⁶Pb/²⁰⁴Pb. Such a variation is quite small, compared with those observed for the North Atlantic as a whole¹⁷ (1.9 for ²⁰⁶Pb/²⁰⁴Pb). No marked difference can be observed between different types of lavas from volcanoes and fissural lava flows.

In the CYAMEX area the strontium isotope values on whole rocks and glasses are uniform and average about 0.7025 which is

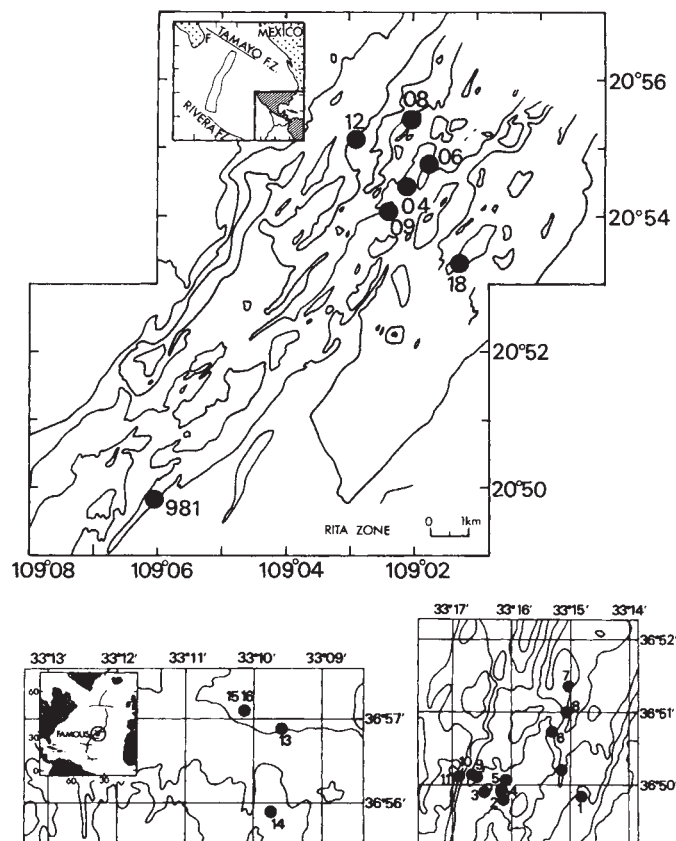


Fig. 1 Location of the FAMOUS and CYAMEX samples.

Table 1 Sr and Pb isotopic compositions of the samples studied in the FAMOUS and CYAMEX areas

FAMOUS Samples	$^{87}\text{Sr}/^{86}\text{Sr}$ WR	$^{87}\text{Sr}/^{86}\text{Sr}$ WR leached	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$	
ARP 74 7-5 C (IG)	0.70292 ± 7	0.70280 ± 10	18.79 ± 0.02	15.53 ± 0.02	38.35 ± 0.05	1
ARP 74 7-5 S (IG)	0.70295 ± 7					
ARP 74 9-12 (IG)			18.69 ± 0.02	15.52 ± 0.02	38.28 ± 0.1	2
ARP 74 9-13 (IG)	0.70295 ± 10	0.70288 ± 6	18.74 ± 0.02	15.52 ± 0.02	38.33 ± 0.05	3
ARP 74 10-14 (IG)		0.70289 ± 8	18.65 ± 0.02	15.53 ± 0.02	38.28 ± 0.05	4
ARP 74 10-15 (IG)			18.83 ± 0.02	15.52 ± 0.02	38.36 ± 0.06	5
ARP 74 11-11 (IG)			18.68 ± 0.02	15.52 ± 0.02	38.20 ± 0.05	6
ARP 74 11-18 (IG)	0.70290 ± 6		18.60 ± 0.02	15.53 ± 0.02	38.23 ± 0.06	7
CYP 30-32 (IG)	0.70302 ± 7	0.70285 ± 6				8
CYP 31-36 (IG)	0.70302 ± 10	0.70285 ± 6				9
		0.70286 ± 6				
CYP 31-37 (IG)	0.70343 ± 17	0.70296 ± 6				10
CYP 31-39 (IG)	0.70315 ± 7	0.70281 ± 7	18.67 ± 0.02	15.50 ± 0.02	38.20 ± 0.07	11
ARP 73 1003 (IG)	0.70287 ± 10	0.70285 ± 7	18.73 ± 0.02	15.53 ± 0.02	38.20 ± 0.07	12
ARP 74 13-21 (FZ)		0.70276 ± 11	18.60 ± 0.02	15.52 ± 0.02	38.20 ± 0.05	13
ARP 74 13-24 (FZ)			18.89 ± 0.02	15.55 ± 0.02	38.45 ± 0.06	14
ARP 74 14-32 (FZ)	0.70309 ± 12	0.70285 ± 8	18.84 ± 0.02	15.52 ± 0.02	38.39 ± 0.08	15
	0.70312 ± 13		18.79 ± 0.02	15.52 ± 0.02	38.31 ± 0.05	
ARP 74 14-33 (FZ)	0.70313 ± 9					16
CH31 DR2 301	0.70315 ± 7					
CH31 DR6 325	0.70329 ± 17		18.54 ± 0.02	15.55 ± 0.02	38.24 ± 0.05	
CH31 DR 10 100	0.70320 ± 8		18.62 ± 0.02	15.54 ± 0.02	38.30 ± 0.07	
CYAMEX						
CYP 78 04 07		0.70254 ± 7	18.33 ± 0.02	15.49 ± 0.02	37.77 ± 0.06	
CYP 78 06 11			18.42 ± 0.02	15.49 ± 0.02	37.85 ± 0.06	
CYP 78 09 12			18.39 ± 0.02	15.51 ± 0.02	37.83 ± 0.07	
CYP 78 12 33			18.37 ± 0.02	15.46 ± 0.02	37.73 ± 0.05	
CYP 78 12 35		0.70250 ± 9				
CYP 78 12 36	0.70307 ± 6	0.70247 ± 7				
CYP 78 18 65		0.70264 ± 8	18.47 ± 0.02	15.49 ± 0.02	37.88 ± 0.06	
981 R10			18.45 ± 0.02	15.49 ± 0.02	37.86 ± 0.06	
981 R23		0.70247 ± 6	18.39 ± 0.02	15.49 ± 0.02	37.83 ± 0.06	

We have divided the samples from the FAMOUS zone into two groups: IG, international ground; FZ fracture zone. The total blank for ~1 g of sample has been maintained below 1 ng for Pb and below 200 pg for Sr. WR, whole rock.

typical of the depleted mid-oceanic ridge basalts (MORB). The lead isotope values are also rather uniform but yield a less radiogenic character than for FAMOUS: $^{206}\text{Pb}/^{204}\text{Pb} = 18.35$, $^{207}\text{Pb}/^{204}\text{Pb} = 15.48$, $^{208}\text{Pb}/^{204}\text{Pb} = 37.85$.

The (Sr, Pb) or (Pb, Pb, Pb) isotope correlation diagrams (Fig. 2) show that both CYAMEX and FAMOUS data plot on the main trend defined for MORB¹⁷⁻¹⁹. Therefore, both zones have quite 'normal' chemical characteristics of an oceanic ridge. The uniformity of their isotopic compositions which contrasts with the variations observed in the North Atlantic^{16,17} and with the local heterogeneities observed in peridotite xenoliths and 'high temperature' peridotites^{20,21} has important implications. In this case, the source of basalts at the scale of 10 km is statistically homogeneous, the formation of a basaltic liquid compensates for any local minor heterogeneity in the mantle source.

Although the Pb values are roughly homogeneous, some small variations in the Pb isotope ratios remain and cannot be ascribed to secondary processes such as seawater alteration. These variations are about one-tenth of the global variation for the North Atlantic (when using the Sr-Pb correlation diagram, we may estimate the corresponding variation for the values obtained close to 0.0001, which for $^{87}\text{Sr}/^{86}\text{Sr}$ is virtually the same as those expected from the analytical error). Thus, the small variations in Pb isotopic ratios are not inconsistent with the uniformity of the Sr isotope ratios. Improvements in measuring the Sr isotopic ratio are necessary before we can study Pb-Sr correlation on such a small scale. However, on this scale these Pb isotope variations are not correlated with variations in ratios of incompatible elements like Hf/Th, Ta/Th, La/Th or La/Sm. Therefore, such small heterogeneity is not an argument against the trace element discussion made by Langmuir *et al.*²² who

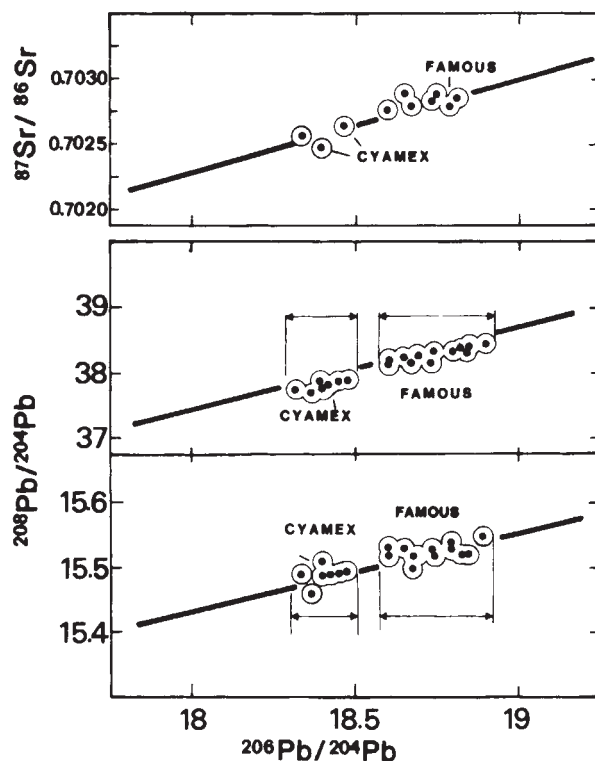


Fig. 2 Results for CYAMEX and FAMOUS areas: a, $^{87}\text{Sr}/^{86}\text{Sr}$ vs $^{206}\text{Pb}/^{204}\text{Pb}$; b, $^{208}\text{Pb}/^{204}\text{Pb}$ vs $^{206}\text{Pb}/^{204}\text{Pb}$; c, $^{207}\text{Pb}/^{204}\text{Pb}$ vs $^{206}\text{Pb}/^{204}\text{Pb}$.

reject the equilibrium partial melting model and propose a new model for partial melting called dynamic-melting.

We now compare the homogeneity observed on a small scale (10 km) with that observed on a larger scale (100–1,000 km).

Results from the North Atlantic^{3,5,17}, and those on the oceanic crust of the Indian Ocean²³ indicate that the cause of isotopic variations obtained along the Atlantic Ridge is linked to mixing between the lower and upper mantle by injection of blobs (or hot spots). Our results show that, on a small scale, the mixing does not generate large isotopic heterogeneities but this conclusion should be evaluated in the light of the type of models discussed in ref. 18. Note that the Pb heterogeneity is larger in the case of FAMOUS than for CYAMEX. Considering that FAMOUS is on a ridge with some blobs mixing while CYAMEX is on a quite 'normal' ridge, such a difference is probably significant. With improvement of the accuracy, it will be quite interesting to study the small variations of Sr isotope ratios in these areas to decipher more precisely the processes of blob mixing and basalt genesis.

This study was supported by CNEXO. Samples were provided by the Centre Océanographique de Bretagne. We thank J. Francheteau and H. Bougault for assistance. This is IGP contribution NS/530.

Received 29 June; accepted 2 October 1981.

1. Tatsumoto, M. *Science* **153**, 1094 (1966).
2. Tatsumoto, M., Hedge, G. E. & Engle, A. G. *J. Science* **150**, 886 (1965).
3. Sun, S. S., Tatsumoto, M. & Schilling, J. G. *Science* **190**, 143 (1975).
4. Hart, S. R., Schilling, J. G. & Powell, J. L. *Science* **246**, 104 (1973).
5. White, W. M. & Schilling, J. G. *Geochim. cosmochim. Acta* **42**, 1501 (1978).
6. Hekinian, R., Moore, J. C. & Bryan, W. B. *Contr. Miner. Petrol.* **58**, 83 (1976).
7. ARCYANA Deep Sea Res. **24**, 565 (1976).
8. Rise Project Group *Science* **207**, 1421 (1980).
9. Juteau, T. *et al. Oceanol. Acta* **3**, 487 (1980).
10. Bougault, H. (in preparation).
11. Francheteau, J. *et al. Nature* **277**, 523 (1979).
12. Birck, J. L. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 37 (1978).
13. Manhès, G., Minster, J. F. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 14 (1978).
14. O'Nions, R. K. & Pankhurst, R. J. *Earth planet. Sci. Lett.* **31**, 255 (1976).
15. White, W. M. *Yb Carnegie Instn Wash.* **325** (1979).
16. White, W. M. & Bryan, W. B. *Bull. geol. Soc. Am.* **88**, 571 (1977).
17. Dupré, B. & Allègre, C. J. *Nature* **286**, 17 (1980).
18. Allègre, C. J., Brevart, O., Dupré, B. & Minster, J. F. *Phil. Trans. R. Soc.* **297**, 447 (1980).
19. Cohen, R. S., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Nature* **283**, 149 (1980).
20. Polve, M. & Allègre, C. J. *Earth planet. Sci. Lett.* **51**, 71 (1980).
21. Dupré, B., Hamelin, B. & Allègre, C. J. *EOS*, **60**, 414 (1979).
22. Langmuir, C. H., Bender, J. F., Bence, A. E., Hanson, G. N. & Taylor, S. R. *Earth planet. Sci. Lett.* **36**, 133 (1977).
23. Dupré, B. & Allègre, C. J. *TERRA Cogn. Proc. 1st EUG Meet. Strasbourg*, 79 (1981).

Palaeoceanographic significance of bottom-current fluctuations in the Southern Ocean

Michael T. Ledbetter

Department of Geology, University of Georgia, Athens, Georgia 30602, USA

Hiatuses in the sedimentary record of the Southern Ocean are usually attributed to erosion by high-velocity bottom currents^{1–3}. To test for palaeoclimatic mechanisms which intensified bottom circulation, however, the timing of increases in bottom-current velocity must be known. As the sedimentary record of the period of initial erosion is lost within the hiatus, the timing of increased velocity must be determined in cores which are adjacent to the axis of highest velocity^{4,5}. In those cores the accumulation rates are reduced due to the current but a record of the period of initiation of scour is preserved as a zone of particle-size winnowing of the fine-fraction^{3–4}. The timing of major episodes of bottom-current intensification in the South Australian Basin for the past 4.5 Myr are reported here for the first time and may be used to examine the role of palaeoclimatic fluctuations on bottom-current intensity.

A series of piston cores across the axis of a high-velocity current (near a modern contour current⁶) which has produced

hiatuses in the sedimentary record² in the South Australian Basin (Fig. 1) are used to demonstrate a method of dating erosional pulses of bottom water. These disconformities are of varying length depending on width, intensity and duration of the high-velocity bottom current responsible for scouring the sea floor. Two of the cores (E48-5 and E45-86) near the axis of flow during an expansion of the contour current have a hiatus which extends from the Miocene to latest Pleistocene while hiatuses in two adjacent cores (E49-53 and E50-2) are shorter (Fig. 2). One core (E45-21) in the area has not recorded a hiatus due to its position at the margin of the bottom water even at its greatest extent during the past 3 Myr. Because three of the cores were the site of deposition during the time span lost due to erosion in the axial-cores (Fig. 2), the particle-size distribution in those cores may be examined for evidence of winnowing which is associated with the erosion.

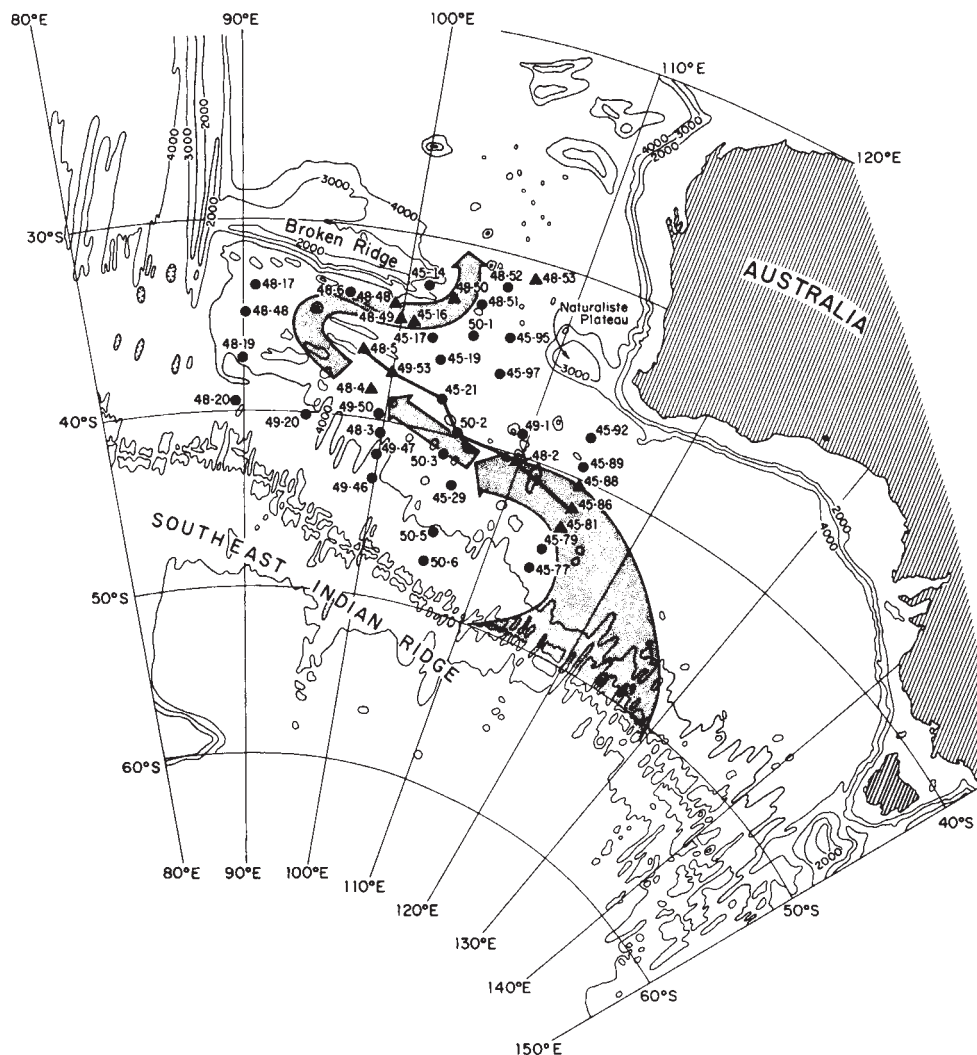
The mean particle-size of the non-carbonate silt fraction (4–62 μm , 8–4 ϕ) was analysed using an electronic particle counter⁷ (Elzone, Particle Data). The carbonate fraction was not analysed as variations in carbonate grain-size may be affected by dissolution of carbonate rather than bottom-current velocity⁸. The silt fraction was analysed because the fine fraction is more sensitive to the magnitude of bottom-current velocity observed in the deep sea^{9,10}. As the cores studied are too far from Antarctica to have a significant ice-rafted debris component¹¹, the grain size of the non-carbonate silt-sized component chiefly reflects bottom-current transport. The mean silt particle size in three cores was plotted against age (Fig. 3) determined from magnetostratigraphy^{1,2}. The mean value of the data in each core is used as an arbitrary reference level to separate periods of relatively high and low inferred bottom-water velocity (Fig. 3).

The period of erosion (or non-deposition) in core E50-2 is characterized by coarse particle sizes in the two cores (E49-53 and E45-21) marginal to the high-velocity zone (Fig. 2). Therefore, the erosional event(s) in E50-2 is 'felt' in the more distant cores by winnowing of the fine fraction. When the period of erosion or non-deposition ceased in core E50-2 the mean particle size deposited was still coarse under the waning current. In the adjacent cores, the mean particle size decreased at the same time that sedimentation resumed above the disconformity. Therefore, the marginal cores have recorded the cessation of high-velocity bottom current at ~2.6 Myr.

The initiation of the high-velocity episode may be dated at the beginning of the winnowing of particles in the marginal cores. While sediment is missing for the period 5.5–4.6 Myr, it is possible to examine the record of bottom-current velocity for the past 4.5 Myr and for a short period at ~5.5–5.6 Myr (Fig. 3). The first episode of high-velocity is marked by a coarsening of the mean particle size at ~4.3 Myr which persisted until ~3.8 Myr. Several shorter pulses with lower magnitude followed the initial coarsening of mean particle size and may record periods of current intensification which either eroded or inhibited deposition at the site of core E50-2. The episodic nature of the inferred bottom-water velocity fluctuations imply that sediment was deposited and subsequently eroded in the scour zone so that the hiatus is a result of multiple high-velocity events.

Unfortunately, the bottom-current velocity record of the lower Gilbert Chron cannot be examined due to the lack of recovery of sediment of that age in the South Australian Basin². If, however, the initial increase in bottom-current velocity occurred at 3.8–4.3 Myr, then ~1 Myr of sediments were eroded at the site of core E50-2 (Fig. 3). Based on average sedimentation rates after the disconformity of 0.35 cm kyr⁻¹ in cores E50-2 and E45-21, the episodes of increased velocity eroded ~3.5 m of sediment deposited before 4.3 Myr and 5 m was eroded and/or not deposited from 4.3 to 2.6 Myr. The total erosional effect of the bottom currents in this area, therefore, is nearly 10 m of sediment. Areas of more intense erosion have occurred in the southeastern Indian Ocean as much older sediment is exposed by other disconformities^{1,2} which must have been nearer the axis of flow.

Fig. 1 Map of South Australian Basin showing major physiographic features, bathymetry and *Eltanin* core locations³. Arrows indicate an inferred bottom-water path based on the presence of cores with hiatuses (▲)². ●, Cores without hiatuses². The pathway is near a modern contour-current identified from benthic foraminiferal assemblages⁶. Profile refers to cores in Fig. 2.



A second unconformity in the area is recorded in core E49-53 (ref. 2). The base of the hiatus is at ~ 1.35 Myr (Fig. 2) and the overlying sediment is known to be < 0.72 Myr because it is all within the normal polarity Brunhes Chron (Fig. 2). Much of the period represented by the hiatus in E49-53 is characterized by high inferred bottom-current velocity in cores E45-21 and E50-2 (Fig. 3). An episode of high velocity is recorded from 1.4 to 0.75 Myr. As the high velocity event began near the time of the base of the hiatus, there may have been no erosion but just winnowing and non-deposition of sediment. Sediment above the disconformity may fall anywhere within the period of low-velocity which encompasses the Brunhes Chron and no attempt was made to determine palaeoveLOCITY fluctuations as no age could be assigned.

The history of AABW velocity may be examined in detail for the first time. The period of increased velocity 3.8–4.3 Myr in the South Australian Basin corresponds to an increased hiatus frequency throughout the south-east Indian Ocean (Fig. 4). The increase in bottom-water velocity in the late Gilbert Chron occurred at the same time as a major cooling in the Antarctic region^{12–14}. Fluctuations observed in the temperature of the sub-Antarctic and Antarctic region¹² closely match the pulses in bottom-water velocity with cold climatic conditions corresponding to increased velocity. When climatic conditions in the sub-Antarctic and Antarctic warmed in the Gauss Chron¹³, the bottom-current velocity waned. The increased benthic circulation with Antarctic region refrigeration occurred due to either increased thermohaline circulation or increased wind stress on the Antarctic circumpolar current; either mechanism could result from steepened Equator to poleward temperature gradients. The Northern Hemisphere refrigeration which resul-

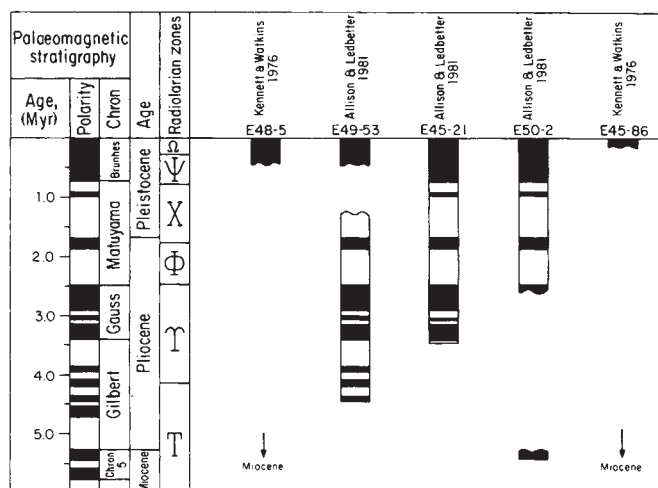


Fig. 2 Magnetostratigraphy of five cores on a profile across the path of a bottom-current in the Australian Basin (see Fig. 1). The disconformities in these cores may be traced into cores E49-53, E45-21 and E50-2 where winnowing of the grain-size population may be used to date the initiation of high velocity events which created a scour zone.

ted in the ice ages, however, occurred at 3.2 Myr (ref. 15) at a time when Antarctic bottom-current circulation waned in the southeastern Indian Ocean (Figs 3, 4), the South Atlantic Ocean⁴ and South Pacific Ocean¹⁶.

The decreased Antarctic-source bottom circulation at the time of the initiation of the Quaternary glaciations may have

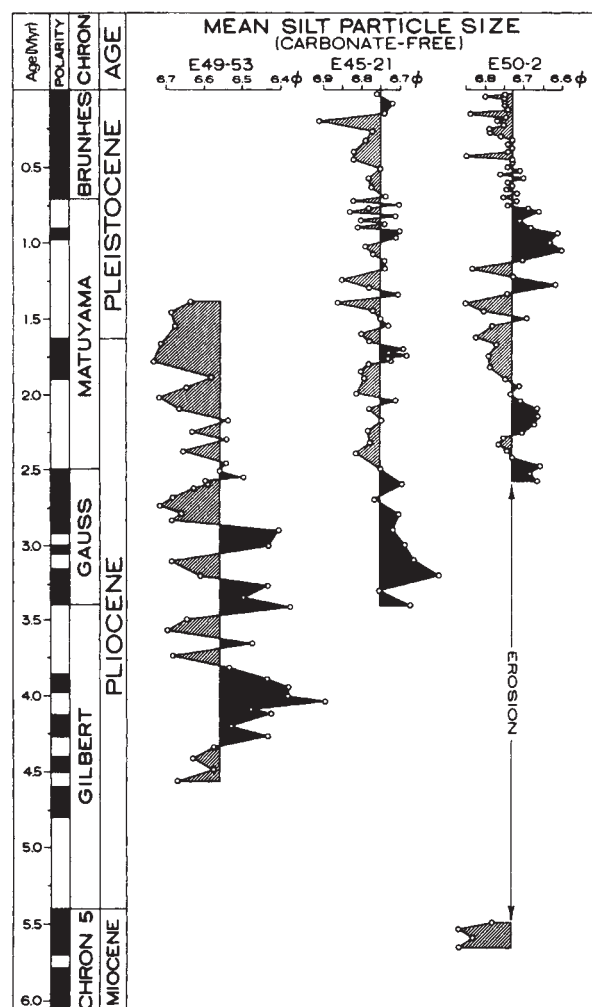


Fig. 3 The mean particle size of the carbonate-free silt fraction in cores E49-53, E45-21 and E50-2 is plotted against age. Blackened areas represent highest relative inferred bottom-water velocity. The disconformity in E50-2 is synchronous with a winnowing zone in E49-53 and E45-21 which is characterized by coarser particle sizes. Highest inferred bottom-water velocity occurred in the late Gilbert and Gauss Chrons. The initiation of winnowing in core E49-53 at 4.3 Myr marks the beginning of the erosion in core E50-2.

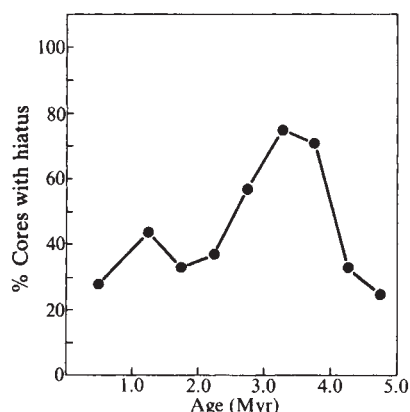


Fig. 4 The percentage of all cores >4440 m in the south-east Indian Ocean^{1,2,18} is ≥ 2.5 –4 Myr BP. The increase in hiatus frequency at ~4 Myr is nearly synchronous with an inferred increase in bottom-current velocity in core E49-53 and the decrease at ~2.5 Myr is nearly synchronous with a decrease in bottom-current velocity inferred in three cores (see Fig. 3).

important palaeoceanographic implications. In the modern ocean, North Atlantic deep water (NADW) provides the salinity component of AABW which increases the density of the cold current to the point where it is the densest water-mass in the ocean. If the production of NADW is affected by the ice ages, then a direct effect may occur in AABW production rates. Two effects may be postulated, however. Cooler Northern Hemisphere climates might increase NADW production through increased sea-ice production in the Norwegian Sea. The cooler climates, however, may result in long periods of thickened sea-ice which halt NADW production by limiting the reservoir for cold saline water within the Norwegian Sea. The latter method has been postulated for Pleistocene glaciations¹⁷ and may be responsible for the decrease of AABW at the onset of the ice ages.

Research was supported by NSF grants DPP79-0511 and DPP81-13147.

Received 18 August; accepted 7 October 1981.

1. Watkins, N. D. & Kennett, J. P. *Ant. Res. Ser.* **19**, 317–334 (1972).
2. Kennett, J. P. & Watkins, N. D. *Bull. geol. Soc. Am.* **87**, 321–339 (1976).
3. Allison, E. T. & Ledbetter, M. T. *Mar. Geol.* (in the press).
4. Ledbetter, M. T., Williams, D. F. & Ellwood, B. B. *Nature* **272**, 237–239 (1978).
5. Huang, T. C. & Watkins, N. D. *Mar. Geol.* **23**, 113–132 (1977).
6. Corliss, B. *Mar. Geol.* **31**, 115–118 (1979).
7. Muerdter, D. R., Dauphin, J. P. & Steele, G. J. *sedim. Petrol.* **51**, 647–650 (1981).
8. Ledbetter, M. T. & Johnson, D. A. *Science* **194**, 837–839 (1976).
9. Ledbetter, M. T. *Mar. Geol.* **33**, 71–89 (1979).
10. Ellwood, B. B. & Ledbetter, M. T. *Earth planet. Sci. Lett.* **35**, 189–198 (1977).
11. Keany, J., Ledbetter, M. T., Watkins, N. D. & Huang, T. C. *Bull. geol. Soc. Am.* **87**, 873–882 (1976).
12. Ciesielski, P. F. & Weaver, F. M. *Geology* **2**, 511–515 (1974).
13. Keany, J. *Mar. Micropaleontol.* **3**, 35–49 (1978).
14. Shackleton, N. J. & Kennett, J. P. *Init. Rep. DSDP Leg 29*, 801–807 (1975).
15. Shackleton, N. J. & Opdyke, N. D. *Nature* **270**, 216–219 (1977).
16. Ledbetter, M. T. & Huang, T. C. *Mar. Geol.* **36**, M21–M28 (1980).
17. Weyl, P. K. *Meteor. Monogr.* **8**, 37–62 (1968).
18. Osborn, N. I. thesis, Univ. Georgia (1981).

C₂₇–C₂₉ ring A monoaromatic steroids in Cretaceous black shales

G. Hussler, B. Chappe, P. Wehrung & P. Albrecht*

Laboratoire de Chimie Organique des Substances Naturelles, associé au CNRS, Institut de Chimie, Université Louis Pasteur, 1 rue Blaise Pascal, 67008 Strasbourg, France

The structural elucidation of aromatic steroid hydrocarbons, which are widespread constituents of sediments and petroleum, has been attempted¹; several series of triaromatic steroids have been identified^{2,3} and a structural hypothesis for the major series of the ubiquitous ring C monoaromatic steroids has been proposed^{2,4}. Although many of these monoaromatic compounds are already present in ancient shales of a low degree of maturity, the stage at which they form and the mechanism of their aromatization are still unclear. Early aromatization of ring A of steroids has been suspected^{5,6}, but never confirmed. Here we report the elucidation of the structures of two series of ring A monoaromatic steroids, 1 and 2, which occur in immature Cretaceous black shales from the Southern and Eastern North Atlantic, collected during the Deep Sea Drilling Project (DSDP). Compounds 1 and 2 were identified as major constituents of the aromatic fractions of these samples by comparison with authentic standards.

Steroid derivatives have been widely used as geochemical markers to study chemical transformations at various stages of diagenesis or maturation in the subsurface. There is good evidence for early alterations of the steroid skeleton in recent or immature ancient sediments: microbial reduction of the Δ^5 double bond, dehydration of stanols leading to Δ^2 -sterenes and the subsequent formation of steranes or isomerization of

* To whom correspondence should be addressed.

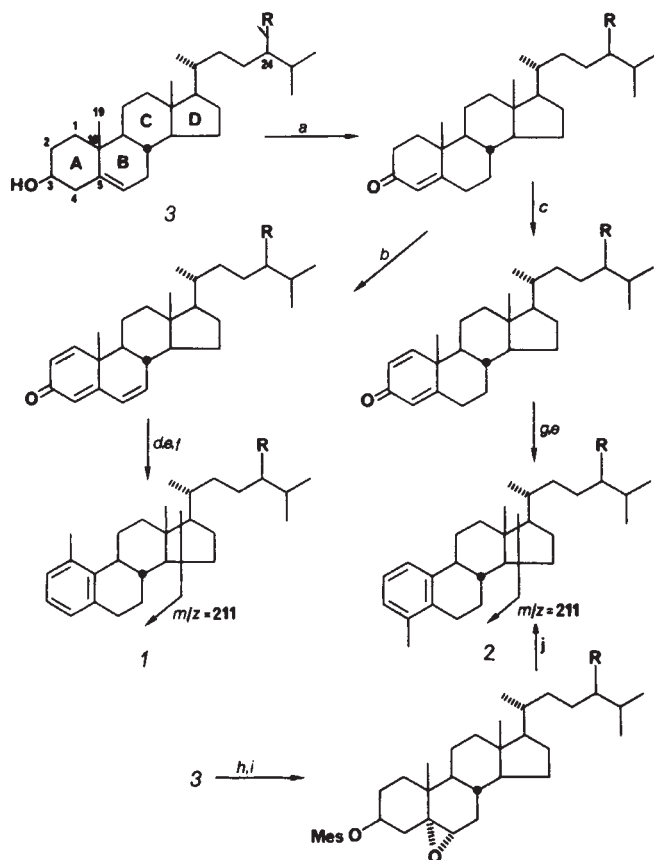


Fig. 1 Scheme of synthesis of ring A monoaromatic steroids, after refs 15–17. *a*, Cyclohexanone, $\text{Al}(\text{O}-i\text{Pr})_3$; *b*, chloranil, pentanol; *c*, DDQ, dioxane; *d*, $i\text{Pr}-\text{OH}$, $\text{Al}(\text{O}-i\text{Pr})_3$; *e*, H^+ ; *f*, H_2 , Pd/C ; *g*, LiAlH_4 ; *h*, $\text{CH}_3\text{SO}_2\text{Cl}$; *i*, $m\text{-ClC}_6\text{H}_4\text{CO}_2\text{H}$; *j*, $\text{CH}_3\text{CO}_2\text{H}$, HBr 48%.

sterenes^{7–11}. However, the aromatization reactions of this class of molecules, which is widespread in living organisms, in particular in phytoplankton, have been little studied¹, despite the fact that the distributions of aromatic steroids are being increasingly used, as a fingerprint, in problems of maturation of geological organic matter³, as well as for correlating crude oils and their source rocks¹².

The black shale samples (10–20 g) were obtained from DSDP Legs 40 (sites 364 and 365) and 41 (site 367) in the Angola and Cape Verde basins. These shales have been deposited in anoxic environments; their organic carbon content is high, ranging from 16 to 5% (refs 13, 14). The samples were extracted with chloroform and total extracts were separated by column and TLC over silica gel. The content of total aromatic hydrocarbons ranged between 17 and 5% of the extract. The mono + diaromatic hydrocarbons were further purified by HPLC (SiO_2 , 10 μm ; heptane) as described previously² and analysed by computerized gas chromatography–mass spectrometry (GC–MS; LKB 9000 S). Steroid hydrocarbons bearing on aromatic A or B ring display a major fragmentation at $m/z = 211$, which can be used for the study of their distribution by mass fragmentography of this specific ion (Fig. 1). The mass fragmentograms of the molecular ions at $m/z = 366$, 380 and 394 show a very similar distribution in which at least four major series are apparent (Fig. 2).

Compounds **1** ($\text{R} = \text{H}$) and **2** ($\text{R} = \text{H}$; $\text{R} = \text{C}_2\text{H}_5$) were prepared from cholest-5-en-3 β -ol (cholesterol; **3**, $\text{R} = \text{H}$) and 24(R)-ethylcholest-5-en-3 β -ol (sitosterol; **3**, $\text{R} = \text{C}_2\text{H}_5$) as a starting material (Fig. 1)^{15–17}. Their structural data agree with those of the literature^{15,16,18} and were confirmed by proton magnetic resonance and mass spectrometry.

1 $\text{R} = \text{H}$ 1-methyl-19-nor-cholesta-1,3,5(10)-triene

NMR (250 MHz; CDCl_3 ; δ p.p.m. JHz): 0.75 (3H,s); 0.87 (6H,d, $J = 6.4$); 0.93 (3H,d, $J = 6.4$); 2.29 (3H,s); 6.69–6.85 (3H,m).

MS (70 eV): $m/z = 366$ (M^+ , 95%); 351 (18%); 253 ($\text{M}^+ - \text{C}_8\text{H}_{17}$, 4%); 226 (17%); 211 (100%); 197 (12%); 183 (12%); 170 (34%); 158 (42%); 157 (64%); 143 (35%); 131 (55%).

2 $\text{R} = \text{H}$ 4-Methyl-19-nor-cholesta-1,3,5(10)-triene

NMR (250 MHz; CDCl_3 ; δ p.p.m., JHz): 0.71 (3H,s); 0.88 (6H,d, $J = 6.4$); 0.95 (3H,d, $J = 6.4$); 2.17 (3H,s); 6.83–7.04 (3H,m).

MS (70 eV): $m/z = 366$ (M^+ , 85%); 351 (2.5%); 253 ($\text{M}^+ - \text{C}_8\text{H}_{17}$, 2%); 226 (15%); 211 (100%); 197 (8%); 183 (6%); 170 (10%); 158 (65%); 157 (32%); 143 (20%); 131 (36%).

$\text{R} = \text{C}_2\text{H}_5$ 4-Methyl-24-ethyl-19-nor-cholesta-1,3,5(10)-triene

NMR (250 MHz; CDCl_3 ; δ p.p.m., JHz): 0.71 (3H,s); 0.83 (3H,d, $J = 6.8$); 0.85 (3H,d, $J = 6.8$); 0.86 (3H,t, $J = 6.8$); 0.95 (3H,d, $J = 6.4$); 2.16 (3H,s); 6.83–7.04 (3H,m).

MS (70 eV): $m/z = 394$ (M^+ , 85%); 379 (2.5%); 253 ($\text{M}^+ - \text{C}_{10}\text{H}_{21}$, 5%); 226 (13%); 211 (100%); 197 (8%); 183 (6%); 170 (10%); 158 (66%); 157 (32%); 143 (16%); 131 (34%).

These compounds were identified with those occurring in the geological samples by comparison of the mass spectra and coelution in single ion mass fragmentography GC–MS on two glass capillary columns (SE 30; SP 2250, Supelco, 25 m $\times$ 0.25 mm). Compounds **1** ($\text{R} = \text{CH}_3$; C_2H_5) and **2** ($\text{R} = \text{CH}_3$) were tentatively identified on the basis of their mass spectra and retention times.

The immature character of these Cretaceous black shales^{13,14} is confirmed by the presence of the less stable $\beta\beta$ -hopane series as major constituents of the alkane and acid fractions, as well as an important fraction of unsaturated hydrocarbons in which monosterenes and hopenes are largely predominant (B.C. and P.A., unpublished data). The identification of monoaromatic steroids **1** and **2** shows that the aromatization of ring A of the precursor sterols begins at a very early stage of diagenesis. These results also demonstrate a novel pathway of aromatization of sterols in the subsurface, starting in ring A rather than in ring C

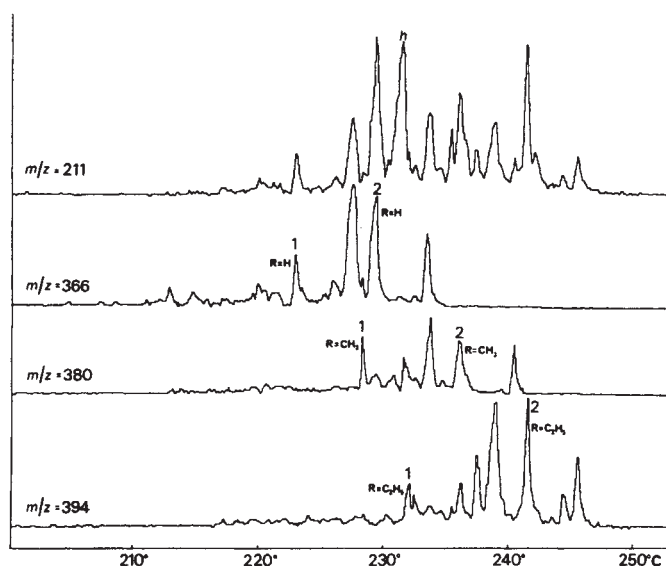


Fig. 2 Mass fragmentograms of ions 211, 366, 380 and 394 showing the distribution of ring A monoaromatic steroid hydrocarbons in a Cretaceous black shale from the DSDP (Leg 40, site 364). Numbers refer to structures in Fig. 1. *h*, Monoaromatic C_{27} hopanoid. Conditions: SP 2250, 25 m $\times$ 0.25 mm, 0.08 μm , 140–250 $^{\circ}\text{C}$, 2 $^{\circ}\text{C min}^{-1}$. Computerized GC–MS, LKB 9000 S.

as previously described^{1,2,4}. Note that only traces of ring C monoaromatic steroids have been detected in our samples which mostly correspond to confined reducing environments. Ring C aromatization obviously needs further isomerization of the precursor sterenes, which seem to be favoured by a more open and oxidizing environment (B.C. and P.A. unpublished data).

The geochemical aromatization reaction proceeds with a shift of the C-19 methyl group of the precursor sterols from position 10 to position 1 or a preferred rearrangement to position 4, both reaction pathways being similar to the action of acids on various unsaturated steroid derivatives in the laboratory^{15,19}. Loss of the angular methyl group seems to be negligible in our case, as indicated by a weak $m/z = 197$ mass fragmentogram. The aromatization process of ring A of sterols could well be due, in its final steps, to an acid catalysis of minerals, clays in particular, acting in the sediments at an early stage of diagenesis. The cholestatrienes leading to the aromatic hydrocarbons 1 and 2 could have been formed by microbial and chemical transformations of the parent sterols^{8,20}. As ring A monoaromatic steroids have not yet been observed in significant amounts in more mature sediments, they are probably intermediates of degradation of steroids restricted to early diagenesis, leading later to triaromatic steroids.

Work is in progress on the identification of the other series of monoaromatic or triunsaturated steroid hydrocarbons occurring in geological samples.

We thank the Institut Francais du Pétrole for financial support and supply of samples, also Professor G. Ourisson for his interest and helpful discussions.

Received 13 July; accepted 6 October 1981.

- Schaeffe, J. thesis, Univ. Louis Pasteur, Strasbourg (1979).
- Ludwig, B., Hussler, G., Wehrung, P. & Albrecht, P. *Tetrahedron Lett.* (in the press).
- Mackenzie, A. S. & Maxwell, J. R. *Geochim. cosmochim. Acta* (in the press).
- Schaeffe, J., Ludwig, B., Albrecht, P. & Ourisson, G. *Tetrahedron Lett.* 4163-4166 (1978).
- Spyckerelle, C. thesis, Univ. Louis Pasteur, Strasbourg (1975).
- Rullkötter, J., von der Dick, H. & Welte, D. H. in *Init. Rep. DSDP Leg 64* (in the press).
- Gaskell, S. J. & Eglinton, G. *Nature* 254, 209-211 (1975).
- Gagosian, R. B. & Farrington, J. W. *Geochim. cosmochim. Acta* 42, 1091-1101 (1978).
- Dastillung, M. & Albrecht, P. *Nature* 269, 678-679 (1977).
- Rubinstein, I., Sieskind, O. & Albrecht, P. *JCS Perkin I* 1833-1835 (1975).
- Ensminger, A., Joly, G. & Albrecht, P. *Tetrahedron Lett.* 1575-1578 (1978).
- Siefert, W. K. & Moldovan, J. M. *Geochim. cosmochim. Acta* 42, 77-92 (1978).
- Herbin, J. P. & Deroo, G. *Docum. Lab. geol. Fac. Sci. Lyon* 75, 71-87 (1979).
- Deroo, G. et al. *Init. Rep. DSDP Leg 41*, 865-873 (1977).
- Dannenberg, H., Neumann, H. G. *Liebigs Ann. Chem.* 646, 148-170 (1961).
- Dannenberg, H., Neumann, H. G. & Dannenberg-von Dresler, D. *Liebigs Ann. Chem.* 674, 152-167 (1964).
- Ogilvie, A. G. & Hanson, J. R. *JCS Perkin I*, 1981-1983 (1972).
- Dannenberg, H. & Gross, H. *J. Tetrahedron* 21, 1611-1617 (1965).
- Bey, P. & Ourisson, G. *Bull. Soc. chim. Fr.* 1402-1410 (1968).
- Charney, W. & Herzog, H. L. *Microbial Transformations of Steroids* (Academic, New York, 1967).

***N*-acetylation is linked to α -MSH release from pars intermedia of the amphibian pituitary gland**

Gerard J. M. Martens, Bruce G. Jenks & A. P. van Overbeek

Department of Zoology, Faculty of Science, Catholic University, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Most amphibians can adapt their colour to that of their background by a mechanism which is controlled by melanophore stimulating hormone (MSH), a product of the pars intermedia of the pituitary gland¹. Although the melanotropic peptides of amphibians have never been sequenced, immunological evidence indicates that at least some of them are structurally related to α -MSH²⁻⁴. In our studies of the biosynthesis and structure of melanotropic peptides in the pars intermedia of *Xenopus laevis*, we have found evidence, reported here, for des-*N*-acetyl- α -MSH as the immediate precursor of α -MSH, with the acetylation taking place just before or during release from the pituitary.

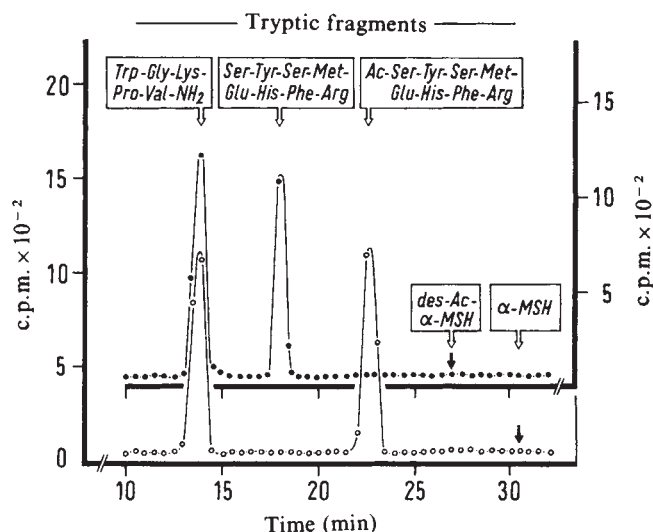


Fig. 1 Tryptic peptide mapping by HPLC of newly synthesized des-*N*-acetyl- α -MSH (●) and α -MSH (○). Neurointermediate lobes of black-adapted *X. laevis* were incubated for 4.5 h at 22 °C in 50 μ l incubation medium (for composition see ref. 5) containing 20 μ Ci each of ³H-lysine, ³H-tyrosine, ³H-phenylalanine and ³H-tryptophan (specific activities 90, 50, 80 and 26 Ci mmol⁻¹, respectively; Amersham). Lobes were extracted with 500 μ l 0.1 M HCl, the extract was analysed with HPLC, and the newly synthesized product co-eluting with synthetic des-*N*-acetyl- α -MSH was isolated from the lobes. To the incubation medium 500 μ l 0.1 M HCl were added, the acidified medium was analysed with HPLC and the newly synthesized product co-eluting with synthetic α -MSH was isolated from the incubation medium. Tryptic digestion of the isolated peptides (to which 20 μ g bovine serum albumin were added) was performed at 37 °C for 2 h with 10 μ g trypsin (DPCC-treated; Sigma) in 10 μ l 55 mM HEPES buffer (pH 8.0); the digest was analysed with HPLC. Digestion of the radioactive tryptic fragments to free amino acids revealed that the 14-min fragments contained radioactive tryptophan and lysine, and both the 18- and 22.5-min fragments contained radioactive tyrosine and phenylalanine. Black arrows indicate the elution positions of the undigested radioactive peptides. For comparison, 50 μ g synthetic des-*N*-acetyl- α -MSH (a gift from Dr G. I. Tesser, Department of Organic Chemistry, Nijmegen) and 50 μ g α -MSH (Peninsula) were digested with 10 μ g trypsin in the same conditions as described for digestion of the radioactive peptides, except no bovine serum albumin was added. The tryptic fragments of the synthetic peptides (shown in upper boxes) were detected with a post-column derivatization system, using *o*-phthalaldehyde (Sigma) as a fluorogenic reagent, and they were further identified with specific staining techniques for the amino acids histidine and tryptophan according to ref. 20. The elution positions of the undigested synthetic peptides are also indicated (lower boxes). HPLC separation was on Spherisorb 10 ODS (Chrompack, 250 $\times$ 4.6 mm i.d.) using a stepwise gradient with a flow rate of 2 ml min⁻¹, and 0.5-min fractions were collected. The gradient was established with 0.5 M formic acid, 0.14 M pyridine (pH 3.0) as the primary solvent and 1-propanol as the secondary solvent.

We have previously shown by *in vitro* pulse-chase techniques coupled to HPLC analysis that the biosynthesis of peptides in the pars intermedia of *X. laevis* begins by the production of a large protein which is subsequently processed to a number of peptides⁵, three of which possess melanotropic activity. One of them has since been shown to co-migrate with synthetic des-*N*-acetyl- α -MSH on acid-urea gels and to have identical retention times to the synthetic peptide during HPLC analysis with several different elution gradients (unpublished observations). Tryptic digestion of this newly synthesized peptide yielded two fragments (Fig. 1) with elution times identical to those of the two fragments obtained from the tryptic digest of synthetic des-*N*-acetyl- α -MSH. Moreover, the radioactive amino acid content of the tryptic fragments was in agreement with that expected on the basis of the amino acid content of the comparable fragments of the synthetic peptide. Therefore, this newly synthesized melanotropic product is des-*N*-acetyl- α -MSH. Another melanotropic peptide co-eluted from the HPLC column with synthetic α -MSH and co-migrated with this peptide during acid-urea gel electrophoresis. The elution times of the two radioactive tryptic fragments of this newly synthesized product (Fig. 1) corresponded exactly with those of the two tryptic

fragments of synthetic α -MSH. Furthermore, the radioactive amino acid content of each tryptic fragment of this product was in full agreement with its identity as α -MSH. The designation of the two newly synthesized products is supported by the observation that the only structural difference between them is located on the N-terminus.

Pulse-chase analysis of the biosynthesis of des-*N* α -acetyl- α -MSH and α -MSH revealed a likely precursor-product relationship for these peptides. During the first 2 h des-*N* α -acetyl- α -MSH accumulated rapidly, then gradually decreased. α -MSH was barely demonstrable in the first 2 h and gradually increased thereafter, concomitant with the decrease of des-*N* α -acetyl- α -MSH (Fig. 2). A further observation was that des-*N* α -acetyl- α -MSH was almost completely restricted to the tissue while α -MSH was exclusively found in the medium. Additional experiments were performed to establish that the absence of des-*N* α -acetyl- α -MSH in the medium was not the consequence of extracellular acetylation or rapid degradation of this peptide. Isolated radioactive des-*N* α -acetyl- α -MSH, incubated with neurointermediate lobes, remained in the non-acetylated form and was stable for at least 8 h. Moreover, superfusion of neurointermediate lobes, whereby any possible modification or proteolysis of secreted peptides was prevented by collecting the perfusate in solutions containing acid or enzyme inhibitors, demonstrated that over 95% of the newly synthesized MSH released was in the acetylated form.

In amphibians, the physiological factor inhibiting MSH release is thought to be dopamine^{6,7}. In *Xenopus*, this catecholamine inhibits the *in vitro* release of all newly synthesized peptides⁸⁻¹⁰. To investigate the effects of inhibition of release on the acetylation process we conducted pulse-chase experiments with dopamine present during the chase. In the control experiment, ~40% of the newly synthesized MSH was in the α -MSH form and, as expected, this acetylated product was found exclusively in the medium (Fig. 3a). The des-*N* α -acetyl- α -MSH was found almost exclusively in the tissue. In the dopamine-treated group, release of newly synthesized peptides was completely inhibited and all MSH in the tissue was des-*N* α -acetyl- α -MSH (Fig. 3b). Thus, inhibition of release was accompanied by inhibition of acetylation.

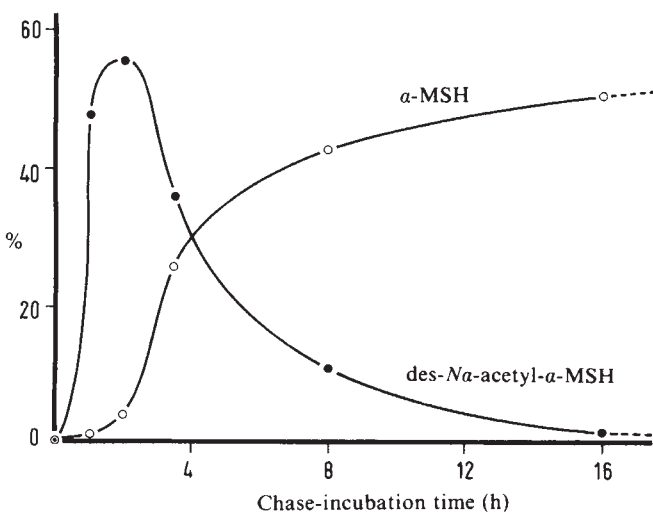


Fig. 2 Levels of newly synthesized des-*N* α -acetyl- α -MSH and α -MSH as a function of chase-incubation time. *Xenopus* neurointermediate lobes were preincubated for 1 h before a 30-min pulse incubation in ^3H -tryptophan, followed by chase incubations in normal incubation medium for 1, 2, 3.5, 8 or 16 h. After the pulse-chase incubations, lobe extracts and incubation media were analysed with HPLC. With the gradient used, newly synthesized des-*N* α -acetyl- α -MSH and α -MSH eluted at 22.5 min and 28 min, respectively. To quantify the relative amounts of these peptides, the chromatographic profile between 20 and 30 min elution time was used. For each chase group, the total level of radioactivity (lobe plus medium) was determined by adding the c.p.m. values of the HPLC fractions. The radioactivity of each HPLC fraction was then expressed as a percentage of this total level. In the figure, the percentages represent the sum of the percentage found in the lobe and that in the medium for des-*N* α -acetyl- α -MSH (●) or α -MSH (○) at each chase-incubation time.

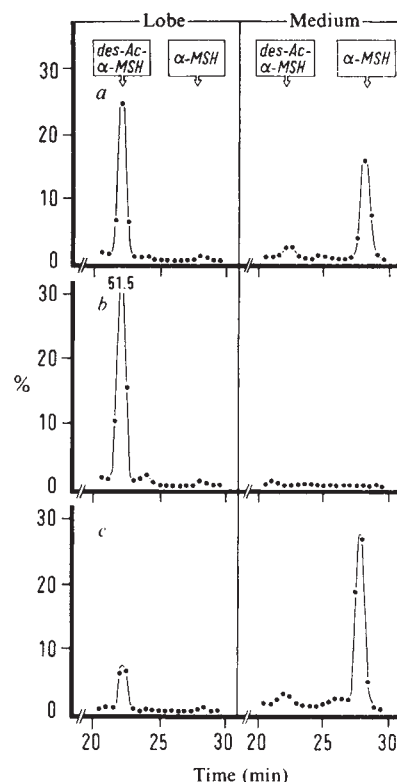


Fig. 3 Effect of dopamine and cyclic AMP on the release of newly synthesized des-*N* α -acetyl- α -MSH and α -MSH. Lobes were preincubated for 1 h before a 30-min pulse with ^3H -tryptophan, followed by chase incubations of 3.5 h in control incubation medium (a), medium containing 10^{-5}M dopamine (3-hydroxytyramine HCl; Sigma) (b), or medium containing 6 mM cyclic AMP (8-bromo cyclic AMP; Sigma) (c). The HPLC fractions are expressed in percentages using the method described in Fig. 2 legend.

Cyclic AMP stimulates release of newly synthesized peptides from the pars intermedia of *Xenopus*¹⁰. The present study shows that cyclic AMP treatment enhanced the amount of α -MSH (Fig. 3c). Approximately 75% of the newly synthesized MSH was present in the α -MSH form and, again, this form was found exclusively in the medium. Thus, stimulation of release was accompanied by stimulation of the acetylation process. Altogether, our results indicate that the acetylation of des-*N* α -acetyl- α -MSH occurs immediately before or during release.

The above results, dealing with newly synthesized peptides, suggest that des-*N* α -acetyl- α -MSH could constitute a 'storage-form' of the hormone. This contention is supported by observations we made concerning stored peptides (unpublished data). Using bioassays, we demonstrated that the pars intermedia of white-adapted *Xenopus*, which is known to store melanotropic peptides³, contains ~10 times more des-*N* α -acetyl- α -MSH than α -MSH. Further, infusing animals adapting to a white background with ^3H -lysine resulted in a preferential accumulation of radioactive des-*N* α -acetyl- α -MSH such that after 10 days >90% of the radioactive MSH in the pars intermedia was in the non-acetylated form.

Des-*N* α -acetyl- α -MSH shows considerably less melanotropic activity than α -MSH¹¹. That acetylation of peptide hormones need not always be accompanied by an increase in activity is, however, illustrated by the fact that β -endorphin is not biologically active in its acetylated form¹². It is interesting that α -MSH and β -endorphin are derived from the same pro-hormone^{13,14}. Whether acetylation constitutes a control point in regulating the biological potency of pars intermedia peptides, as suggested earlier¹², remains to be established.

In the rat pituitary gland, Woodford and Dixon¹⁵ demonstrated the presence of *N* α -acetyltransferase, which can acetylate the α -amino group of synthetic des-*N* α -acetyl- α -MSH to form α -MSH. The acetylation reaction reported in the present investigation is probably catalysed by such an enzyme. As this

acetylation is closely related to the release process, the enzyme is probably associated with the secretory granules or with the cell membrane. This situation would be unique, as most acetylations of polypeptides take place before or during packaging in the secretory granules^{12,16}; a pertinent example is the α -acetylation of β -endorphin in the rat pars intermedia¹⁷.

The process of exocytosis implies the fusion of the membrane of the secretion granule with the plasma membrane¹⁸. Membrane fusion, in general, may be accompanied by physical and chemical changes in the membranes involved¹⁹. An intriguing possibility for the acetylation of des- α -acetyl- α -MSH in the amphibian pars intermedia is that during the secretory process physicochemical changes in the microenvironment near the cell membrane lead to activation of a specific α -acetyltransferase, thus linking acetylation of the hormone to its release.

This work was supported by a grant from the Netherlands Organization for the Advancement of Pure Research (ZWO) through the Foundation for Biological Research (BION).

Received 15 May; accepted 14 October 1981.

1. Lerner, A. B., Shizume, K. & Bunding, I. *J. clin. Endocr. Metab.* **14**, 1463–1490 (1954).
2. Shapiro, M. et al. *Endocrinology* **90**, 249–256 (1972).
3. Jenks, B. G., Van Overbeek, A. P. & McStay, B. F. *Can. J. Zool.* **55**, 922–927 (1977).
4. Vaudry, H. et al. *Gen. comp. Endocr.* **34**, 391–401 (1978).
5. Martens, G. J. M., Jenks, B. G. & Van Overbeek, A. P. *Comp. Biochem. Physiol.* **67**, 493–497 (1980).
6. Hadley, Mac. E. & Hruby, V. J. *Am. Zool.* **17**, 809–821 (1977).
7. Jenks, B. G. in *Frontiers of Hormone Research* (ed. Van Wimersma Greidanus, Tj. B.) 63–65 (Karger, Basel, 1977).
8. Loh, Y. P. & Gainer, H. J. *gen. Physiol.* **70**, 37–58 (1977).
9. Jenks, B. G. & Van Overbeek, A. P. *Comp. Biochem. Physiol.* **66**, 71–76 (1980).
10. Martens, G. J. M., Jenks, B. G. & Van Overbeek, A. P. *Comp. Biochem. Physiol.* **69**, 75–82 (1981).
11. Schwyzer, R. & Eberle, A. in *Frontiers of Hormone Research* (ed. Van Wimersma Greidanus, Tj. B.) 18–25 (Karger, Basel, 1977).
12. Smyth, D. G., Massey, D. E., Zakarian, S. & Finnie, M. D. A. *Nature* **279**, 252–254 (1979).
13. Mains, R. E., Eipper, B. A. & Ling, N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5300–5304 (1977).
14. Roberts, J. L. & Herbert, E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4826–4830 (1977).
15. Woodford, T. A. & Dixon, J. E. *J. biol. Chem.* **254**, 4993–4999 (1979).
16. Strous, G. J. A. M., Berns, A. J. M. & Bloemendal, H. *Biochem. biophys. Res. Commun.* **58**, 876–884 (1974).
17. Seizinger, B. R. & Höllt, V. *Biochem. biophys. Res. Commun.* **96**, 535–543 (1980).
18. Douglas, W. W. *Biochem. Soc. Symp.* **39**, 1–28 (1974).
19. Papahadjopoulos, D., Protis, A., Pangborn, W. & Newton, C. in *Transport of Macromolecules in Cellular Systems* (ed. Silverstein, S. C.) 413–430 (Dahlem Konferenzen, Berlin, 1978).
20. Block, R. J., Durrum, E. L. & Zweig, G. (eds) *A Manual of Paper Chromatography and Paper Electrophoresis* 2nd edn, 133–137 (Academic, New York, 1958).

The regulatory subunit of adenylate cyclase interacts with cytoskeletal components

Mark M. Rasenick*, Peter J. Stein & Mark W. Bitensky

Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

The adenylate cyclase system, which mediates cellular responses to a variety of hormones and neurotransmitters, contains at least three plasma membrane-associated proteins: the hormone receptor, the regulatory or guanyl nucleotide-binding unit (G unit) and the catalytic moiety. Activation of the enzyme after binding of hormone to the receptor involves binding of GTP to the G unit¹. Activation can be enhanced by increasing the fluidity of the membrane with unsaturated fatty acids² with the local anaesthetic prilocaine³, and perhaps, *in vivo*, by phospholipid methylation⁴. Disruption of microtubules increases localized mobility of membrane proteins in a manner similar to that of unsaturated fatty acids⁵ and can enhance cyclic AMP accumulation in intact leukocytes^{6–9}. We therefore decided to investigate whether there is a direct interaction between microtubules and adenylate cyclase. We show here that colchicine, vinblastine and *cis*-unsaturated fatty acids enhance G unit-mediated activation of adenylate cyclase, implying that microtubules or tubulin are involved in the attachment of the G unit to the membrane.

* Author to whom correspondence should be addressed at: Department of Neurology, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

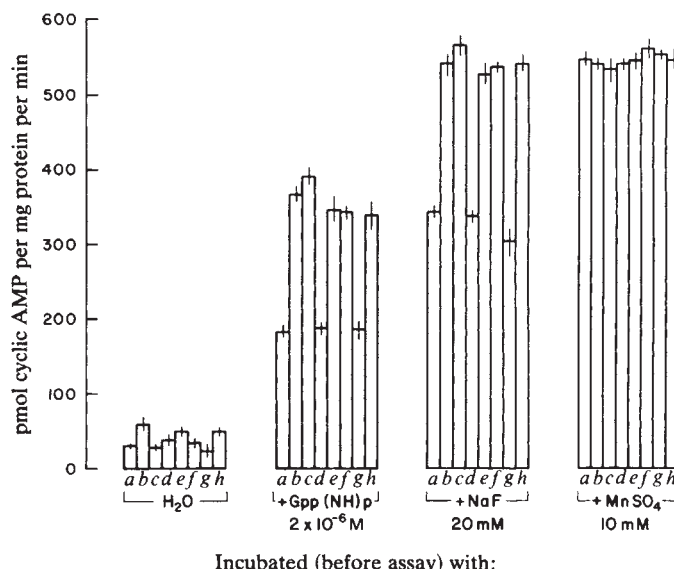


Fig. 1 Adenylate cyclase activity in synaptic membranes preincubated with fatty acids or microtubule-disrupting agents. Enriched synaptic membrane fractions (ESMF) were prepared from 21-day old male Sprague-Dawley rats by a modification²⁵ of the method of Gray and Whittaker²⁶. After washing the 10,000 g pellet three times in a 0.32 M sucrose, 20 mM Tris-maleate pH 7.8, 5 mM MgSO₄, 2 mM EGTA, 1 mM dithiothreitol (DTT) buffer, the membranes (about 80–85% synaptosomes by electron microscopy) were stored under liquid nitrogen. After thawing, synaptosomes were lysed in 20 mM Tris-maleate pH 7.8, 5 mM MgSO₄, 1 mM DDT (1.2 mg protein per ml) and treated at 30 °C, 20 min with compounds a–h as indicated. 40 µl of the membrane suspension were then incubated (30 °C, 20 min) with 10 µl of Gpp(NH)p, NaF, Mn²⁺ or H₂O. Adenylate cyclase assays²⁵ are started by the addition of ATP and a regenerating system (100 µl total volume), carried out for 10 min at 30 °C and stopped by boiling. Cyclic AMP produced is assayed by protein binding²⁷ using a rabbit skeletal muscle-binding protein²⁸. Values are the means of triplicate determinations in one of four similar experiments. a, H₂O; b, colchicine 10⁻⁶ M; c, vinblastine 10⁻⁶ M; d, lumicolchicine 10⁻⁵ M; e, *cis*-vaccenic acid 10 µg ml⁻¹; f, linoleic acid 10 µg ml⁻¹; g, stearic acid 10 µg ml⁻¹; h, linoleic acid + colchicine.

In enriched synaptic membrane fractions (ESMF) from rat cerebral cortex, the G unit (with GTP) can interact directly with the catalytic moiety to activate adenylate cyclase. In this case, the hormone receptor is functionally bypassed^{10,11} and the interaction of two, rather than three, subunits can be examined while the regulatory role of the G unit is maintained. Thus when fluoride or guanylimidodiphosphate (Gpp(NH)p—a hydrolysis-resistant GTP analogue) is present, adenylate cyclase is stably activated by the G unit–catalytic moiety complex. Manganese, however, can bypass the G unit to activate the catalytic moiety directly¹².

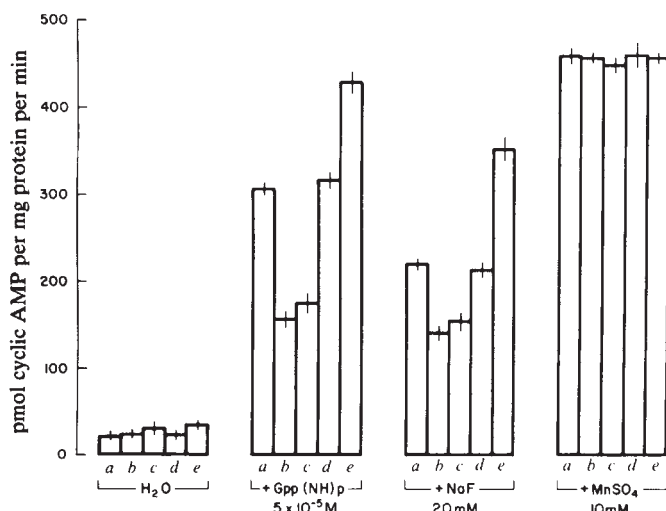
In an attempt to increase the G unit–catalytic moiety interaction by increasing the mobility of the proteins, ESMF were incubated with colchicine (K_a 5 × 10⁻⁷ M) or vinblastine and the effects of these agents on Gpp(NH)p or fluoride activation of adenylate cyclase was examined. In these conditions, adenylate cyclase activity is significantly increased (Fig. 1). Lumicolchicine, which does not bind to tubulin¹³, does not have this effect at any concentration. Incubation with membrane fluidizers such as *cis*-vaccenic or linoleic acid is approximately as effective as colchicine or vinblastine in elevating Gpp(NH)p or NaF activation of adenylate cyclase. The effects of fatty acids and colchicine are not additive. When fatty acids, colchicine or vinblastine were added, they did not alter the catalytic activity as determined by activity in the presence of manganese.

Membranes that have been incubated with colchicine or vinblastine, and subsequently washed, show a loss of Gpp(NH)p or NaF-mediated adenylate cyclase activity. Treatment of ESMF with *cis*-vaccenic acid, however, continues to enhance adenylate cyclase activity even after the membranes have been washed. Again the intrinsic catalytic activity, as measured with manganese, was not altered by these treatments (Fig. 2).

To test whether any material that may have been stripped from the membrane during colchicine preincubation would restore or enhance Gpp(NH)p and fluoride sensitivity, we added back concentrated, dialysed supernatants from the washes following preincubation with colchicine. We found that both the control supernatant (data not shown) and, to a greater extent, the colchicine preincubation supernatant (dialysed free of colchicine and guanyl nucleotide) augmented adenylate cyclase activity (both basal and in the presence of Gpp(NH)p or fluoride: Fig. 3). These data suggest that the G unit may be liberated from the synaptic membrane.

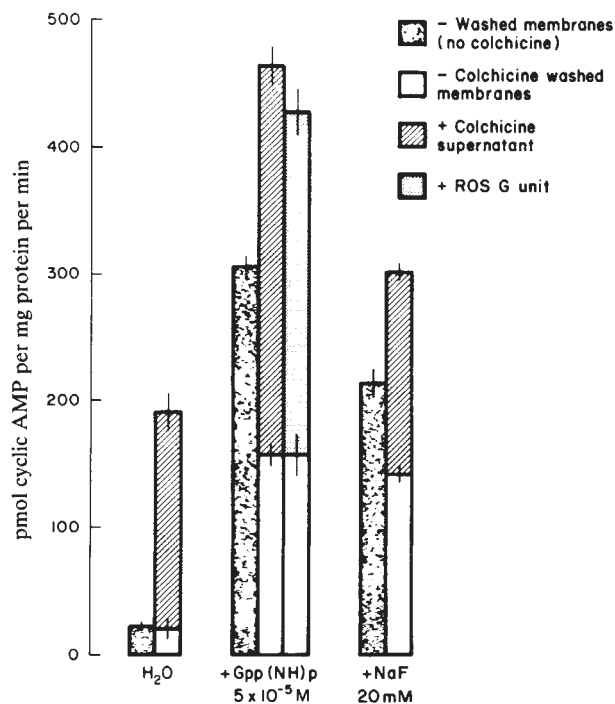
Previous experiments in this laboratory have indicated that the GTPase complex (G unit) from amphibian or bovine photoreceptor rod outer segment cyclic GMP phosphodiesterase (ROS-PDE) is similar to the G unit of adenylate cyclase¹⁴, and that ESMF adenylate cyclase can be activated by photoreceptor components¹⁵. The ROS G unit becomes a stable activator of ROS-PDE or adenylate cyclase when Gpp(NH)p is bound in place of GTP^{16,17}. Thus, when purified preparations of ROS G unit are added to ESMF which have been depleted of guanyl nucleotide or fluoride responsiveness by colchicine, Gpp(NH)p and NaF responsiveness is restored. The degree of restoration is comparable with that observed when concentrated supernatants from membranes preincubated with colchicine and subsequently washed (colchicine/wash supernatant) are added to ESMF (Fig. 3).

While the ability to add back Gpp(NH)p and fluoride responsiveness to colchicine-treated ESMF gives a qualitative indica-



Adenylate cyclase assayed in the presence of:

Fig. 2 Preincubation of synaptic membranes with microtubule-disrupting agents or fatty acids followed by washing: effects on adenylate cyclase. ESMF were treated (as in Fig. 1) with the indicated compounds and then washed twice with 10 volumes of 10 mM Tris-maleate pH 7.8, 2 mM EDTA, 1 mM DTT followed by centrifugation at 10,000g for 15 min (4°C). Membranes were resuspended and assayed for adenylate cyclase in the presence of H₂O, Gpp(NH)p, NaF or MnSO₄ as noted for 10 min at 30°C. Values are means of triplicate determinations for one of three similar experiments. Fluoride activation is lower than that shown in Fig. 1, as NaF is present only during the assay step²⁵. a, No addition; b, colchicine 5×10^{-6} M; c, vinblastine 5×10^{-6} M; d, lumicolchicine 10^{-5} M; e, *cis*-vaccenic acid $10 \mu\text{g ml}^{-1}$.



Adenylate cyclase assayed in the presence of:

Fig. 3 Reconstitution of adenylate cyclase activity with soluble components. Supernatants were prepared by incubating ESMF from 12 cerebral cortices with 10^{-5} M Gpp(NH)p ($\pm 5 \times 10^{-6}$ M colchicine) for 20 min at 30°C and washing with 10 mM Tris-maleate pH 7.8, 2 mM EDTA, 1 mM DTT. The process was repeated twice and the three 10,000g supernatants pooled and centrifuged for 1 h at 105,000g. The resulting supernatant was concentrated and dialysed free of Gpp(NH)p or colchicine in an Amicon Cf 25 cone (molecular weight cut-off $\sim 25,000$) in 20 mM Tris-maleate pH 7.8, 5 mM MgSO₄, 1 mM DTT. 10 μl of a 0.74 mg protein per ml solution were added where indicated. ROS G unit was prepared (as described elsewhere²⁹) by exposing toad (*Bufo marinus*) ROS to light and 2×10^{-6} M Gpp(NH)p. The ROS G unit consists of three polypeptides with apparent molecular weights of 39,000, 37,000 and 10,000, and these three species comprise >99% of the material added. The supernatant was concentrated (as above for ESMF) and 10 μl of a 0.12 mg protein per ml solution were added where indicated. Both the ROS G unit and the ESMF supernatants are devoid of adenylate cyclase and PDE activity, and do not alter the PDE activity of the ESMF.

tion that G unit activity can be lost and replaced, it is necessary to quantify the replacement. Towards this end, endogenous G unit activity was depleted with differential heat inactivation¹⁸. When ESMF are heated at 42°C for 30 min, about 15% of Gpp(NH)p or NaF-mediated adenylate cyclase activity is retained compared with about 54% of catalytic activity (as monitored with Mn²⁺). Colchicine/wash supernatant or ROS G unit are capable of restoring the Gpp(NH)p and NaF response of heated membranes to a level concomitant with that of remaining catalytic activity (Fig. 4).

Activation of adenylate cyclase by ROS G unit and colchicine/wash supernatant saturates at 0.012 and 0.074 mg protein per ml respectively. Note that higher amounts of purified ROS G than endogenous G unit are required to reconstitute adenylate cyclase activity within the ESMF. This may be because ROS G is less able to insert successfully into the adenylate cyclase syncytium than is the native G unit. Nonetheless, it appears that the maximal Gpp(NH)p or NaF responsiveness conferred by added ROS G unit or colchicine/wash supernatant is limited by the Mn²⁺-determined catalytic activity (data not shown).

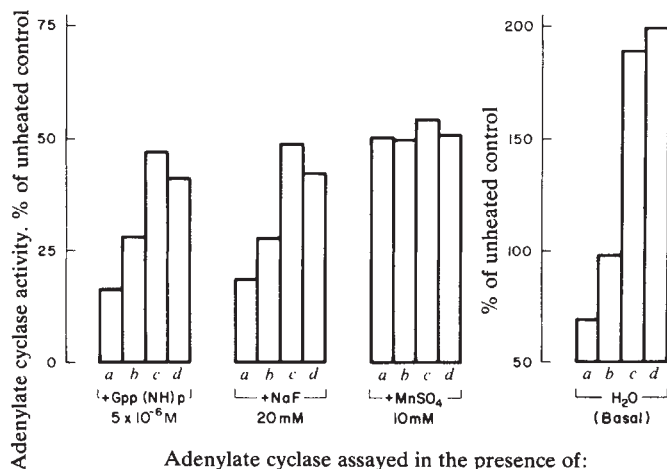


Fig. 4 Restoration of partially heat-inactivated adenylate cyclase activity with soluble synaptic membrane and rod outer segment components. ESMF were incubated at 42 °C for 30 min before assay for adenylate cyclase activity. Synaptic membrane supernatant ($\pm$ colchicine) or ROS G unit (prepared as described in Fig. 3 legend) were added to the ESMF. The membranes were then assayed for adenylate cyclase (as described in Fig. 3 legend). Values are means of triplicate determinations from one of four similar experiments and are expressed as % of unheated controls. Unheated control values are: H₂O, 43 pmol cyclic AMP per mg protein per min; Gpp(NH)p (5×10^{-6} M), 290; NaF (20 mM), 231; MgSO₄ (10 mM), 547. Supernatant additions: a, none; b, +control (no colchicine); c, +colchicine; d, +ROS G unit.

Figure 4 shows that Gpp(NH)p cannot reconstitute G unit-mediated adenylate cyclase activation without added ROS G unit or synaptic membrane supernatant; increasing Gpp(NH)p (beyond 5×10^{-6} M) has no effect on adenylate cyclase activity. The supernatants, however, are capable of conferring additional guanyl nucleotide stimulation. ROS G unit or colchicine/wash supernatant charged with GDP do not increase basal adenylate cyclase activity in the differentially heat inactivated membranes (data not shown). When colchicine/wash supernatant or ROS G unit are heated (90 °C, 5 min) and added back to differentially heat-inactivated synaptic membranes, adenylate cyclase activities do not differ significantly from control (no supernatant added) values. Thus, these data indicate not only that free nucleotide is not responsible for restoration of adenylate cyclase activity, but that a heat-labile macromolecule with bound Gpp(NH)p is required.

The data presented here demonstrate that both free fatty acids and microtubule-disrupting agents enhance activation of rat synaptic membrane adenylate cyclase. Membrane-fluidizing lipids^{2,19} or the local anaesthetic prilocaine³, and colchicine or vinblastine⁶⁻⁹ are reported to activate adenylate cyclase in a number of cellular systems, and this activation seems to be distal to the hormone receptors^{2,3,8}. The interaction of vinblastine at the G unit has been suggested to explain the biphasic effects of that compound in rat liver plasma membranes²⁰. Our data suggest that the locus of the fatty acid and microtubule disruptor stimulation of synaptic membrane cyclase is at the G unit, as manganese response is not affected by these agents.

It is possible that the colchicine-mediated loss of Gpp(NH)p or fluoride responsiveness is due to a release of the G unit from the membrane. The ability to add back the same amount of G unit activity with colchicine/wash supernatant as with photo-receptor G unit supports this idea.

Most attempts to isolate the G unit have used detergents¹². However, low salt washes can dislodge some G unit from the pigeon erythrocyte membrane²¹, there may be a soluble or a loosely attached G unit in liver cells²² and previous experiments have observed a soluble factor which augments GTP or hormone sensitivity in liver plasma membranes²³. The data

presented here indicate that a subset of (though not necessarily all) G units may have a colchicine sensitive attachment to the synaptic plasma membrane.

The facilitation of G unit-mediated adenylate cyclase activity with either free fatty acids or microtubule-disrupting agents (without washing) suggests that the ability of the G unit to diffuse laterally in the membrane is a limiting factor in cyclase activation; that is, agents known to increase freedom of protein mobility in the membrane consequently increase cyclase activation. This effect may occur as a result of an increase in whole membrane fluidity or as a local effect at the site of G unit anchoring. Linoleic and *cis*-vaccenic (but not stearic) acids increased adenylate cyclase activity in these studies. The two former compounds (but not the latter) have recently been demonstrated to block lymphocyte capping in a manner similar to colchicine. In this sense, the fatty acids and microtubule disruptors might release constraints and enhance membrane protein mobility within localized domains⁵. A localized increase in membrane fluidity has similarly been proposed to explain the prilocaine augmentation of fluoride activated hepatic adenylate cyclase³.

Loss of G unit activity mediated by colchicine or vinblastine but not fatty acids may be explained if one hypothesizes a microtubule (or membrane tubulin²⁴) attachment site. We propose that fatty acids and microtubule-disrupting agents activate adenylate cyclase by different mechanisms. *cis*-Unsaturated fatty acids may increase membrane protein mobility, increasing the probability of effective G unit interaction with a catalytic moiety. Colchicine or vinblastine, however, may free the G unit from an attachment site, thus facilitating G unit-catalytic moiety interaction (and loss of G unit after washing). As not all the G unit activity can be eluted by colchicine, even after repetitive washing, it is likely that not all G units have a microtubule or tubulin attachment site. These data support a possible cytoskeletal-adenylate cyclase interaction and suggest additional possibilities for the control of adenylate cyclase activity.

We thank Drs H. Malech, J. Morrow and D. U'Prichard for critical reading of this manuscript and Dr R. Sorbi for helpful discussions. This work is supported by NIH grant AM 20179. M.M.R. is the recipient of a NIH postdoctoral Fellowship AM 06287.

Received 4 May; accepted 6 October 1981.

- Rodbell, M. *Nature* **284**, 17-22 (1980).
- Hanski, E., Rimon, G. & Levitzki, A. *Biochemistry* **18**(5), 846-853 (1979).
- Houslay, M. D. *et al. Biochem. J.* **190**, 131-137 (1980).
- Hirata, F., Strittmatter, W. J. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 368-372 (1979).
- Klausner, R. D., Bhalla, D. K., Dragsten, P., Hoover, R. L. & Karnovsky, J. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 437-441 (1980).
- Rudolph, S. A., Greengard, P. & Malawista, S. E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3404-3408 (1977).
- Rudolph, S. A., Hegstrand, L. F., Greengard, P. & Malawista, S. E. *Molec. Pharmac.* **16**, 805-812 (1979).
- Insel, P. A. & Kennedy, M. S. *Nature* **273**, 471-473 (1978).
- Kennedy, M. S. & Insel, P. A. *Molec. Pharmac.* **16**, 215-223 (1979).
- Iversen, L. *Neurochemistry* **29**, 5-12 (1977).
- Daly, J. in *Cyclic Nucleotides in the Nervous System* (Plenum, New York, 1977).
- Ross, E. M. & Gilman, A. G. *Rev. Biochem. Phys.* **53**, 533-564 (1980).
- Wilson, L. & Friedkin, M. *Biochemistry* **5**, 2463-2468 (1966).
- Shinozawa, T., Sen, I., Wheeler, G. L. & Bitensky, M. W. *J. supramolec. Struct.* **10**, 185-190 (1979).
- Bitensky, M. W., Wheeler, G. L., Yamazaki, A., Rasenick, M. M. & Stein, P. in *Molecular Mechanisms of Photoreceptor Transduction* (ed. Miller, W. H.) (Academic, New York, in the press).
- Fung, B. K. K., Hurley, J. B. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 152-156 (1981).
- Uchida, S., Wheeler, G. L., Yamazaki, A. & Bitensky, M. W. *J. cyclic Nucleotide Res.* **7**, 95-104 (1981).
- Brostrom, M., Brostrom, C., Breckenridge, B. & Wolfe, D. *Adv. Cyclic Nucleotide Res.* **9**, 85-99 (1978).
- Colard, O., Kervabon, A. & Roy, C. *Biochem. biophys. Res. Commun.* **95**, 97-102 (1980).
- Whetton, A. D. & Houslay, M. D. *FEBS Lett.* **111**, 290-293 (1979).
- Pfeuffer, T. *J. biol. Chem.* **252**, 7224-7234 (1977).
- Krishna Bhat, M., Iyengar, R., Abramowitz, J., Bordelon-Riser, M. E. & Birnbaumer, L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3836-3840 (1980).
- Pecker, F. & Hanoune, J. *J. biol. Chem.* **252**, 2784-2786 (1977).
- Bhattacharyya, B. & Wolff, J. *Nature* **264**, 576-577 (1976).
- Rasenick, M. M. & Bitensky, M. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4628-4632 (1980).
- Gray, E. & Whittaker, V. J. *Anat. Rec.* **96**, 79-88 (1962).
- Brown, B., Ekins, R. & Albano, J. *Adv. Cyclic Nucleotide Res.* **2**, 25-40 (1972).
- Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **67**, 305-312 (1970).
- Kuhn, H. *Nature* **283**, 587-589 (1980).

A single amino acid change in IFN- β_1 abolishes its antiviral activity

H. Michael Shepard*, David Leung*,
Nowell Stebbing† & David V. Goeddel*

Departments of *Molecular Biology and †Biology, Genentech Inc.,
460 Point San Bruno Boulevard, South San Francisco,
California 94080, USA

The cloning and expression of human fibroblast interferon (IFN- β_1) cDNA sequences in *Escherichia coli* have been reported by several groups¹⁻³; this will allow extensive study of its biological and physical properties. Aside from the cloning and expression of interesting and useful cDNA sequences, recombinant DNA techniques make possible the *in vitro* construction of gene sequences encoding novel polypeptides which do not occur naturally^{4,5}, which is difficult to achieve by direct chemical modification of proteins. We describe here the isolation of an IFN- β_1 cDNA sequence encoding the 117 C-terminal amino acids of a variant IFN- β_1 (Cys→Tyr at amino acid position 141) and *in vitro* recombination with sequences coding for the N-terminal amino acids of IFN- β_1 to form a DNA sequence encoding a mature variant IFN- β_1 polypeptide, and its expression in *E. coli*. We show that the replacement of Cys at position 141 with Tyr inactivates antiviral activity in a cytopathic effect (CPE) inhibition assay and that the variant protein does not compete with normal fibroblast interferon for reaction with heterospecific antibodies to IFN- β_1 . These effects are attributed to an alteration in protein structure of IFN- β_1 ^{Cys→Tyr} which result from its inability to form an essential disulphide bond.

In our efforts to isolate a recombinant plasmid containing the cDNA sequences encoding human fibroblast interferon (IFN- β_1), several partial-length cDNA clones were obtained³. Among these is pF526, which is ~640 base pairs (bp) long and contains the poly(A) sequence, the 3'-untranslated region and about two-thirds of the coding region for the mature IFN- β_1 protein. DNA sequence analysis showed two changes from the published sequence of IFN- β_1 (Fig. 1; refs 1-3). The first involves the deletion of nucleotides 762-764 (GTC) immediately preceding the poly(A) in the 3'-noncoding region and may be due to normal variation in processing/polyadenylation of the primary transcript. The second change is a G→A transition at position 485 of the published sequence, which lies within the coding region for the mature polypeptide and results in the replacement of a cysteine (TGT) codon with a tyrosine (TAT) codon. Because all the leukocyte interferons (IFN- α) described to date⁶ have a cysteine residue which can be aligned with the cysteine at position 141 of IFN- β_1 and as this cysteine is involved in disulphide bond formation in the IFN- α family⁷, it seemed likely that IFN- β_1 ^{Cys→Tyr} might be a useful variant for investigating the structure of the IFN- β_1 molecule in relation to its antiviral activity.

As the cDNA insert of pF526 comprises only a partial copy of the IFN- β_1 mRNA and encodes only the 117 C-terminal amino acids of the mature polypeptide (Fig. 1), it was necessary to prepare a hybrid gene containing DNA sequences for the N-terminal 56 amino acids of the mature polypeptide and pF526 sequences encoding the remaining 110 acids. As shown in Fig. 2, pFIF *trp*³ 69 (ref. 3) is a pBR322 derivative which carries three *trp* promoter fragments all oriented so that transcription occurs towards the cDNA insert encoding mature IFN- β_1 . The IFN- β_1 cDNA expression insert is bounded by *Xba*I (5') and *Bgl*II (3') restriction endonuclease sites³. We removed the *Xba*I-*Bgl*II insert from pFIF *trp*³ 69, discarded the 3' *Hga*I-*Bgl*II gene fragment and replaced it with the corresponding sequences from the variant. The recombined sequences were inserted into the original expression vehicle containing three promoters to

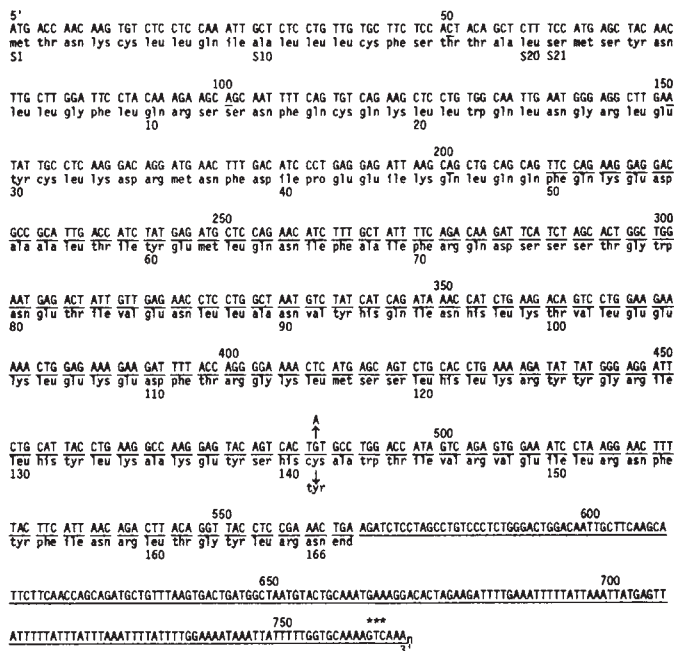


Fig. 1 The normal and variant IFN- β_1 cDNA sequences. Numbers above each line refer to nucleotide position, and numbers below refer to amino acid residue (S, signal peptide). The variant cDNA clone is underlined. The G→A transition is indicated by an arrow at position 485. The three nucleotides deleted in pF526 are indicated by asterisks above the normal sequence at positions 762-764. These cDNAs were cloned and isolated as described elsewhere³.

achieve maximal expression (Fig. 2). Restriction endonuclease and DNA sequence analysis of the new expression plasmid, pFIF *trp*³ 202, showed that the promoters, the ribosome binding site and the entire hybrid coding region gene were arranged correctly (results not shown).

The interferon activity present in lysates of *E. coli* 294 containing the wild-type (pFIF *trp*³ 69) and variant (pFIF *trp*³ 202) expression plasmids as measured in a CPE inhibition assay³, was 1.6×10^4 U ml⁻¹ for 294/pFIF *trp*³ 69 extracts but <21 U ml⁻¹ for extracts prepared from 294/pFIF *trp*³ 202 (Table 1). Because the Cys→Tyr alteration may have affected the folding of the polypeptide, rendering it more labile to *E. coli* proteases, it was necessary to show that a stable protein with some of the properties of IFN- β_1 was detectable in extracts prepared from *E. coli*. As biosynthetic fibroblast interferon comprises $<0.1\%$ of *E. coli* protein it was not possible to show that equivalent amounts of both proteins were present in the extracts used for the initial assays shown in Table 1. To confirm that IFN- β_1 ^{Cys→Tyr} was inactive we synthesized both proteins in a coupled *in vitro* transcription-translation system derived from *E. coli* PR13 (ref. 8). Both pFIF *trp*³ 69 and pFIF *trp*³ 202 were used to programme the *in vitro* system and the products were analysed by SDS-polyacrylamide gel electrophoresis (Fig. 3) and tested for antiviral activity. Equivalent amounts of both wild-type and variant proteins were synthesized *in vitro* (Fig. 3).

Table 1 Antiviral activities of wild-type and variant IFN- β_1

Sample	Dilution of lysate	Interferon activity (U ml ⁻¹)
FIFN <i>trp</i> ³ 69	—	1.6×10^4
FIFN <i>trp</i> ³ 69	10 ⁻¹	1.3×10^3
FIFN <i>trp</i> ³ 69	10 ⁻²	1.7×10^2
FIFN <i>trp</i> ³ 202	—	<21
FIFN <i>trp</i> ³ 202	10 ⁻¹	<21
FIFN <i>trp</i> ³ 202	10 ⁻²	<21

Lysates were prepared and assayed by CPE inhibition³. Different dilutions are shown to indicate the range of variability in the assay.

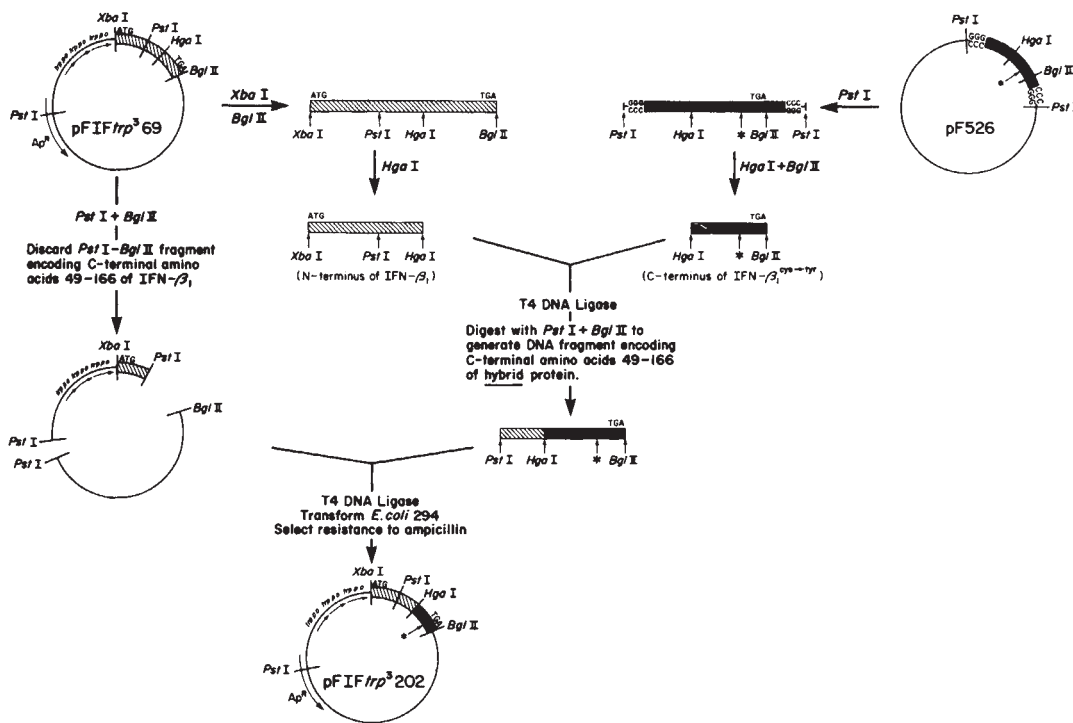


Fig. 2 Construction of the IFN- β_1 ^{Cys→Tyr} expression plasmid. The starting materials were constructed as described elsewhere³. All enzymes were from New England Biolabs and digestions were done in 60 mM Tris pH 7.5, 6 mM MgCl₂, 60 mM NaCl and 6 mM 2-mercaptoethanol at 37 °C. Ligations were done overnight at 12 °C in 60 mM Tris pH 7.5, 3 mM MgCl₂ and 10 mM dithiothreitol. Transformations were done as described previously³. Transformed cells were selected on LB-agar plates containing 20 μ g ml⁻¹ ampicillin to ensure that only clones containing plasmids with the correct structure were recovered. Arrows within plasmids represent *trp* promoter fragments and their direction of transcription. Cross-hatched blocks represent wild-type coding sequences and shaded blocks variant coding sequences.

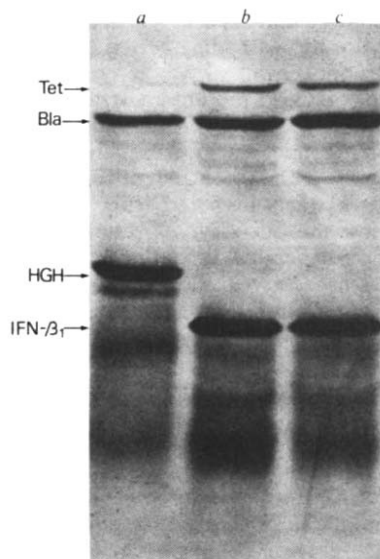


Fig. 3 *In vitro* synthesis of normal and variant IFN- β_1 . Plasmids encoding human growth hormone (a), IFN- β_1 ^{Cys→Tyr} (b) and wild-type IFN- β_1 (c) were used to programme a coupled *in vitro* transcription-translation system. The top two bands in each case are the plasmid-encoded protein for tetracycline resistance (Tet; molecular weight (*M_r*) 36,000) and β -lactamase (Bla; *M_r* 32,000). The lower band in a is human growth hormone (HGH; *M_r* 22,000) encoded by a plasmid similar in structure to the expression plasmids encoding the variant and wild-type IFN- β_1 (*M_r* 20,000). The *in vitro* system used was that described in ref. 8. Plasmid DNA (5 μ g) was added to a total reaction volume of 25 μ l. 5 μ l of each reaction mixture was analysed by electrophoresis through a 15% SDS-polyacrylamide gel¹² and subsequently fluorographed as described by Bonner and Laskey¹³. 10 μ l of each reaction mixture was used to determine antiviral activity in a CPE inhibition assay³.

When aliquots of the *in vitro* reactions were used in a CPE inhibition assay, extract programmed by pFIF *trp*³ 69 yielded 3×10^3 U ml⁻¹ IFN- β_1 antiviral activity, whereas pFIF *trp*³ 202 extracts had no detectable activity.

Although the variant fibroblast interferon, IFN- β_1 ^{Cys→Tyr} shows no antiviral activity it might still compete with the other form of IFN- β_1 in binding antibodies against fibroblast interferon. Incubation of IFN- β_1 with up to a 1,000-fold excess of IFN- β_1 ^{Cys→Tyr} and sufficient heterospecific rabbit anti-fibroblast interferon antibodies (obtained from the Antiviral Substances Program of the National Institutes of Allergy and Infectious Diseases) to neutralize the wild-type IFN- β_1 only, no antiviral activity was observed (data not shown). These results indicate that the amino acid change in IFN- β_1 ^{Cys→Tyr} disrupts the structure of the molecule sufficiently to prevent it from binding antibody specific for that part of the protein essential for induction of antiviral activity in target cells.

The Cys → Tyr change introduced in the fibroblast interferon variant described here could disrupt the function of the molecule in a number of ways, for example, by altering the strength of intramolecular hydrogen bonding or hydrophobic interactions. However, because preincubation of IFN- β_1 extracts with dithiothreitol completely abolishes antiviral activity in a CPE inhibition assay (H.M.S., unpublished observations), it is likely that the disulphide bond between Cys 31 and Cys 141 is essential to maintain the structure of IFN- β_1 required for antiviral activity. Our data therefore suggest that the altered structure of IFN- β_1 ^{Cys→Tyr} prevents it from binding to cell membrane receptors, an interaction which mediates the antiviral activity of interferon⁹.

Variant interferons have been found in the leukocyte interferon family, notably two forms of LeIFN H, which could represent allelic forms⁶. Between the IFN- α species A and D there are a total of 29 amino acid differences⁶. Of these, three occur within the region which is most conserved among the IFN- α species (between amino acids 115 and 151; refs 6, 10). Therefore, some amino acid sequence changes within this conserved region are tolerated, although the cysteine at position 139 (141 of IFN- β_1) may be essential for antiviral activity

(R. Wetzel, personal communication). Because the 5'-terminal sequences of IFN- β_1 ^{Cys→Tyr} are unknown it is uncertain whether it is analogous to the variants of the leukocyte interferon family. It is possible that the G→A transition at nucleotide 486 (Fig. 1) is the result of a reverse transcription error and does not occur naturally. Of four partial-length cDNA clones that were obtained from the human fibroblast cell line GM-2504A (ref. 3), only pF526 had the Cys→Tyr change at amino acid 141 and the deletion of nucleotides 762–764 immediately preceding the polyadenylation site. In any case it is possible that the variant IFN- β_1 mRNA identified by cDNA cloning is not normally translated, or if it is, it could be biologically active by virtue of some compensating changes in the unknown segment of the gene. The variant IFN- β_1 gene cannot correspond to the proposed second fibroblast interferon gene (IFN- β_2) because IFN- β_2 sequences show little or no cross-hybridization with IFN- β_1 (ref. 11). Cell membrane binding studies are in progress to determine directly whether IFN- β_1 ^{Cys→Tyr} is unable to bind cell receptors.

Received 28 July; accepted 14 October 1981.

1. Taniguchi, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 5230–5233 (1980).
2. Derynck, R. *et al. Nature* **285**, 542–547 (1980).
3. Goeddel, D. V. *et al. Nucleic Acids Res.* **8**, 4057–4074 (1980).
4. Streuli, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2848–2852 (1981).
5. Weck, P. K. *et al. Nucleic Acids Res.* (submitted).
6. Goeddel, D. V. *et al. Nature* **290**, 20–26 (1981).
7. Wetzel, R. *Nature* **289**, 606–607 (1981).
8. Brawner, M. & Jascunas, S. R. *J. molec. Biol.* (in the press).
9. Stewart, W. E. II *The Interferon System* (Springer, New York, 1979).
10. Streuli, M., Nagata, S. & Weissman, C. *Science* **209**, 1343–1347 (1980).
11. Weissenbach, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7152–7156 (1980).
12. Laemmli, U. K. *Nature* **277**, 680–685 (1970).
13. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83–88 (1974).

Non-muscle α -actinins are calcium-sensitive actin-binding proteins

Keith Burridge* & James R. Feramisco

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

Many actin-binding proteins have been purified from different cell types, and a number of them share certain features: they are affected by calcium in their interaction with actin and have similar subunit molecular weights (~100,000) on SDS-polyacrylamide gels. These proteins have been given various names, including gelsolin^{1–3}, villin^{4,5}, actinogelin^{6,7}, *Dictyostelium* 95K protein^{8,9}, platelet 95K protein¹⁰ and *Acanthamoeba* 85K protein¹¹. The actin-binding protein α -actinin has received little attention in this analysis, even though it has a subunit molecular weight in this range (~100,000) and has been shown immunologically to exist in non-muscle cells^{12,13}. Here we report that α -actinin purified from HeLa cells is inhibited from cross-linking actin filaments by micromolar free calcium. In many properties it closely resembles actinogelin, *Dictyostelium* 95K protein and *Acanthamoeba* 85K protein, but it appears distinct from proteins such as gelsolin and villin which, in the presence of calcium, fragment actin filaments into short oligomers^{3,5}. We conclude that in this molecular weight range there are at least two classes of non-muscle α -actinin-binding protein: one corresponding to non-muscle α -actinins and the other to proteins such as gelsolin and villin.

α -Actinin has been purified from cultured HeLa cells using a modification of the procedure we developed for the purification of α -actinin from smooth muscle¹⁴. Chromatography of this α -actinin on hydroxyapatite resulted in the separation of two

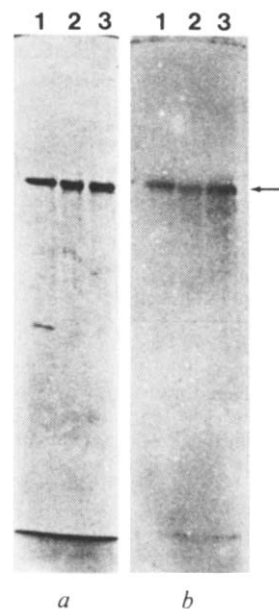


Fig. 1 SDS gel analysis of HeLa cell α -actinins and α -actinins from chicken gizzard smooth muscle. *a*, A 10% polyacrylamide SDS gel stained with Coomassie blue on which were electrophoresed 4 μ g of the following proteins: track 1, smooth muscle α -actinin; track 2, HeLa α -actinin I; track 3, HeLa α -actinin II. The slight difference in molecular weight between the two forms of HeLa cell α -actinin is not detectable at this level of protein loading on the gel. *b*, An autoradiograph of the same gel slice after staining the gel with antibodies against beef cardiac α -actinin followed by a second radio-iodinated antibody directed against the first antibody. Note that the antibody against α -actinin labels all three proteins. This antibody has been characterized previously¹⁸ and only a tight doublet of bands corresponding to the two α -actinins is labelled when a gel of total HeLa cell proteins is stained with this antibody¹⁸. The gel staining with antibody was performed as described elsewhere¹⁹. SDS-gel electrophoresis was performed using Laemmli's buffer conditions²⁰.

forms with slightly different subunit molecular weights on SDS gels (~102,000 and 100,000). The detailed description of the purification and characterization of these two forms of HeLa α -actinin will be described elsewhere (J.R.F. and K.B., in preparation). Figure 1 shows SDS-polyacrylamide electrophoretic gels of the two forms (referred to here as HeLa α -actinin I and HeLa α -actinin II) next to a sample of smooth muscle α -actinin (Fig. 1*a*). An autoradiograph of the same gel is also shown (Fig. 1*b*) after incubation with antibody against muscle α -actinin followed by incubation with a second radio-iodinated antibody directed against the first antibody. This indicates that both HeLa proteins cross-react immunologically with muscle α -actinin. In addition to their similar purification properties, subunit molecular weights on SDS gels and cross-reaction with antibodies to muscle α -actinin, these two HeLa proteins have been identified as α -actinin by several criteria that will be described in detail elsewhere (J.R.F. and K.B., in preparation). Briefly, both native proteins are dimers, with Stokes' radii and *s* values close to those of muscle α -actinin. Their native peptide maps and amino acid composition are similar to but distinct from those of muscle α -actinin. Both proteins bind to F-actin and this binding is diminished by tropomyosin at 37 °C, as is found with muscle α -actinin. Unlike muscle α -actinin, however, the binding to F-actin is markedly inhibited by low free calcium concentrations (Figs 2, 3).

In low free calcium (<10⁻⁷ M), the two non-muscle α -actinins are potent actin gelation factors. The concentration of the α -actinins that will cause gelation of F-actin (as determined by the failure of a stainless-steel ball to fall in the falling ball viscometer) varies between experiments and is dependent on conditions such as actin concentration, temperature and age of the α -actinin preparations. Typically, with F-actin at

* Present address: Laboratories for Cell Biology, Department of Anatomy, University of North Carolina, Chapel Hill, North Carolina 27514, USA.

Table 1 Classification of calcium-sensitive actin-binding proteins

I	Non-muscle α -actinin class
	HeLa α -actinins
	Platelet 105K protein
	Actinogelin
	<i>Dictyostelium</i> 95K protein
II	<i>Acanthamoeba</i> 85K protein
	Gelsolin/villin class
	Gelsolin
	Villin
	Platelet 95K protein

The class I proteins cross-link F-actin at low calcium concentrations but are inhibited from cross-linking at high calcium concentrations. They appear to be rod-shaped proteins, and in some cases have been shown to be dimeric. The class II proteins are globular, monomeric proteins that in the presence of calcium fragment actin filaments into short oligomers.

0.2 mg ml⁻¹ gelation is seen at an α -actinin: actin molar ratio of 1:100–1:200, but at 0.4 mg ml⁻¹ F-actin gelation has been seen at a molar ratio of about 1:400 and this ratio will decrease still further as the actin concentration is increased.

The effect of different free calcium concentrations on the viscosity of mixtures of actin and α -actinin is shown in Fig. 3. Smooth muscle α -actinin is essentially unaffected by calcium concentrations in the range examined here. With the two forms of HeLa α -actinin a sharp fall in the viscosity occurs when the free calcium concentration exceeds 10⁻⁷ M. Note, however, that the viscosity only falls to the background level of the F-actin on its own, indicating that the cross-linking by α -actinin has been inhibited. On the other hand, with a protein such as gelsolin in the presence of micromolar calcium and F-actin, the viscosity of the mixture decreases below the level of the F-actin control due

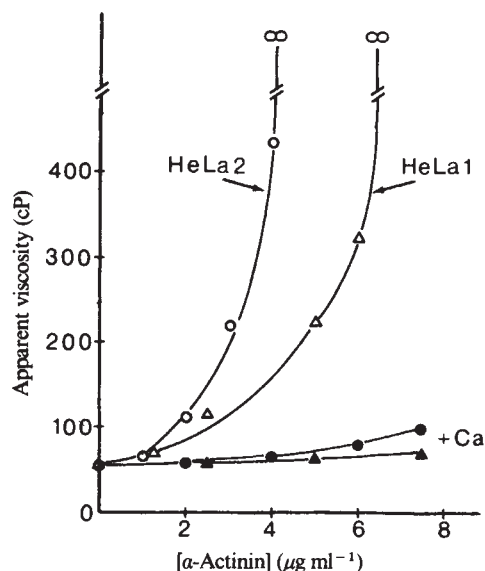


Fig. 2 Effect of varying concentration of HeLa α -actinin on the viscosity of F-actin. The two forms of HeLa α -actinin were added to preformed F-actin (0.2 mg ml⁻¹) at low (10⁻⁷ M) and high (10⁻⁴ M) free calcium. Δ , HeLa α -actinin I at low free calcium; $\blacktriangle$, HeLa α -actinin I at high free calcium; $\circ$, HeLa α -actinin II at low free calcium; $\bullet$, HeLa α -actinin II at high free calcium. The final ionic conditions were 40 mM KCl, 1 mM MgCl₂, 0.1 mM ATP, 0.275 mM EGTA, 0.1% β -mercaptoethanol, 10 mM PIPES, pH 7.0, for those measurements at <10⁻⁷ M free calcium, or for those measurements at 10⁻⁴ M free calcium the same ionic conditions were used except that the EGTA was replaced by 0.1 mM CaCl₂. Apparent viscosity was measured using the low-shear, falling ball viscometer²¹ at 21 °C, 2 h after mixing the samples. Note that in the presence of 10⁻⁴ M calcium the viscosity remained essentially at the background level generated by F-actin alone (apparent viscosity = 56 cP), whereas below 10⁻⁷ M calcium gelation was induced by ~5 μ g ml⁻¹ of HeLa α -actinin II or 7.5 μ g ml⁻¹ of HeLa α -actinin I.

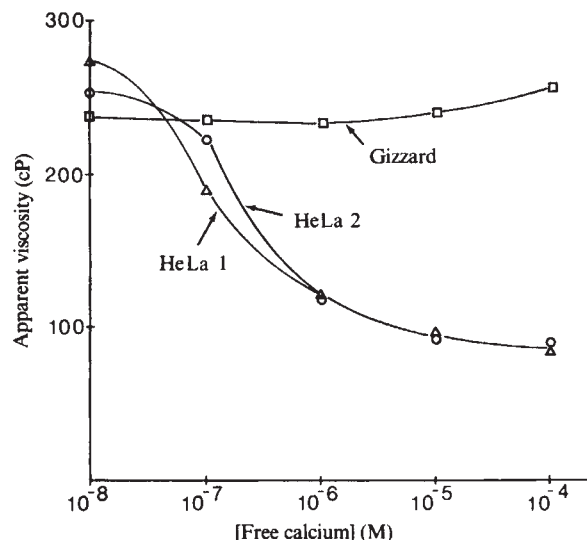


Fig. 3 Effect of varying free calcium concentration on the viscosity generated by mixing purified HeLa cell or chicken gizzard smooth muscle α -actinins with F-actin. HeLa α -actinin I (Δ) was used at a final concentration of 5 μ g ml⁻¹, HeLa α -actinin II ($\circ$) was used at 3.5 μ g ml⁻¹, chicken gizzard smooth muscle α -actinin ($\square$) was used at ~10 μ g ml⁻¹. The α -actinins were added to preformed F-actin, used at a final concentration of 0.2 mg ml⁻¹. The final ionic conditions were identical to those in Fig. 2 except the free calcium concentrations were generated using a Ca-EGTA buffer in which the EGTA concentration was maintained at 0.275 mM, apart from measurements of the values at 10⁻⁴ M calcium, from which EGTA was omitted and replaced by 0.1 mM CaCl₂. Free calcium concentrations were calculated using an apparent association constant for EGTA and calcium of log 6.68 (ref. 22). Any effect of 0.1 mM ATP was ignored in this calculation. Viscosity was measured using a falling ball viscometer at 23 °C 3 h after mixing the samples. The apparent viscosity of the F-actin without α -actinin was 78 cP and was unaffected by the different calcium concentrations. Note that in the higher free calcium concentrations the viscosity of the samples with either of the two HeLa α -actinins fell to a level approaching that of the F-actin alone but did not fall below this value.

to the severing of the actin filaments to short oligomers by this protein³.

The two forms of α -actinin from HeLa cells are very similar to the muscle α -actinins by all criteria so far examined, with the exception of this calcium-inhibited interaction with actin. In collaboration with Rosenberg and Stracher, we have recently demonstrated that a platelet protein of molecular weight 105,000 (105K protein) that is calcium sensitive in its binding to actin is the platelet α -actinin¹⁵. The calcium sensitivity of both HeLa and platelet α -actinins suggests that this is a general property of non-muscle α -actinins. On examining the other actin-binding proteins in this molecular weight range, we note that several closely resemble the HeLa α -actinins in response to calcium and in many of their other properties. Actinogelin closely resembles muscle α -actinin in all physical properties examined so far (A. Asano, personal communication). *Acanthamoeba* 85K protein resembles α -actinin in being a rod-shaped protein with a Stokes' radius of about 85 Å (ref. 11). *Dictyostelium* 95K protein is also rod-shaped and preliminary data indicate that, like α -actinin, it is dimeric (D. L. Taylor, personal communication). All these proteins cross-link F-actin in a manner that is inhibited by calcium^{7,8,11}. On the other hand, proteins such as gelsolin and villin seem to be quite different, being globular, monomeric proteins that also bind to actin but, in the presence of calcium, fragment actin filaments to shorter oligomers^{2,3,5}. We have, therefore, grouped these actin-binding proteins in Table 1 into two distinct classes: an α -actinin class which would include the HeLa α -actinins of the present study, the platelet 105K protein, actinogelin, *Dictyostelium* 95K protein and *Acanthamoeba* 85K protein; and a second class that would include gelsolin and villin. As T. P. Stossel and his

colleagues have identified gelsolin-like proteins immunologically in a wide variety of different vertebrate cell types (personal communication), it may be that most higher eukaryotic cells contain representatives of both classes of protein. So far, a protein similar to gelsolin has not been identified in this molecular weight range in lower eukaryotes, and the equivalent function may be served by a protein of lower molecular weight. Such a protein, called fragmin, with an apparent molecular weight of 45,000, has been purified from *Physarum*^{16,17}, and a low molecular weight protein with similar properties has also been identified in *Dictyostelium* (J. Spudich, personal communication).

The difference between non-muscle α -actinins and muscle α -actinins in their calcium-sensitive interaction with actin may reflect the different requirements of the muscle and non-muscle contractile systems. With muscle, a comparatively stable actin

lattice is required for efficient repeated contraction, but with most non-muscle cells movement is generally accompanied by a change of shape. It seems necessary that the cell's force-generating apparatus should be able to be disassembled and reorganized rapidly for movement to continue. Given this requirement, it may well be an advantage for a non-muscle cell to have an α -actinin that will respond to a calcium flux. In the absence of calcium, non-muscle α -actinin would cross-link actin filaments to form a stable network, but in the presence of calcium it would release the actin filaments and so permit their rapid reorganization.

We thank J. D. Watson for support and encouragement, F. Matsumura for help with calculating the free calcium concentrations and V. Fowler for advice on the falling ball viscometer. This work was supported by NIH grants GM 27781 and GM 28277.

Received 13 July; accepted 18 September 1981.

1. Yin, H. L. & Stossel, T. P. *Nature* **281**, 583–586 (1979).
2. Yin, H. L. & Stossel, T. P. *J. biol. Chem.* **255**, 9490–9493 (1980).
3. Yin, H. L., Zaner, K. S. & Stossel, T. P. *J. biol. Chem.* **255**, 9494–9500 (1980).
4. Bretscher, A. & Weber, K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2321–2325 (1978).
5. Bretscher, A. & Weber, K. *Cell* **20**, 839–847 (1980).
6. Mimura, N. & Asano, A. *Nature* **272**, 273–276 (1978).
7. Mimura, N. & Asano, A. *Nature* **282**, 44–48 (1979).
8. Hellewell, S. B. & Taylor, D. L. *J. Cell Biol.* **83**, 633–648 (1979).
9. Condeelis, J. in *International Cell Biology* (ed. Sweiger, H. G.) 306–320 (Springer, Berlin, 1981).
10. Wang, L.-L. & Bryan, J. *J. Cell Biol.* **87**, 224a (1980).
11. Pollard, T. D. *J. biol. Chem.* **256**, 7666–7670 (1981).
12. Lazarides, E. & Burridge, K. *Cell* **6**, 289–298 (1975).
13. Scholmeyer, J. V., Furcht, L. T., Goll, D. E., Robson, R. M. & Stromer, M. M. in *Cell Motility* (eds Goldman, R., Pollard, T. D. & Rosenbaum, J.) 361–388 (Cold Spring Harbor Laboratory, New York, 1976).
14. Feramisco, J. R. & Burridge, K. *J. biol. Chem.* **255**, 1194–1201 (1980).
15. Rosenberg, S., Stracher, A. & Burridge, K. *J. biol. Chem.* (in the press).
16. Hasegawa, T., Takahashi, S., Hayashi, H. & Hatano, S. *Biochemistry* **19**, 2677–2683 (1980).
17. Hinssen, H. in *Cell Motility: Molecules and Organization* (eds Hatano, S., Ishikawa, H. & Sato, H.) 59–85 (University of Tokyo Press, 1979).
18. Burridge, K. & McCullough, L. *J. supramolec. Struct.* **13**, 53–65 (1980).
19. Burridge, K. *Meth. Enzym.* **50**, 54–63 (1978).
20. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
21. MacLean-Fletcher, S. D. & Pollard, T. D. *J. Cell Biol.* **85**, 414–428 (1980).
22. Amos, W. B., Routledge, L. M., Weis-Fogh, T. & Yew, F. F. *Soc. exp. Biol.* **30**, 273–301 (1976).

Myosin filaments have non-phosphorylated light chains in relaxed smooth muscle

A. V. Somlyo, T. M. Butler*, M. Bond & A. P. Somlyo

Pennsylvania Muscle Institute, University of Pennsylvania, School of Medicine, B42 Anatomy–Chemistry Building/G3, Philadelphia, Pennsylvania 19104, USA

* Thomas Jefferson University, Department of Physiology, 1020 Locust Street, Philadelphia, Pennsylvania 19107, USA

The provocative hypothesis that myosin in resting smooth muscle is in a soluble form and assembles into filaments only at the time of contraction^{1,2} has been disproved with the demonstration of regular arrays of myosin filaments in a variety of smooth muscles, including those shown to have been relaxed before and during fixation^{3–7}. Nevertheless, following the observations of Suzuki *et al.*⁸ on the effect of light-chain phosphorylation on smooth muscle myosin assembly *in vitro*, this hypothesis has been revived with the speculation that relaxed smooth muscle contains non-phosphorylated, non-filamentous myosin that assembles into filaments only after the muscles are stimulated and myosin is phosphorylated⁹. Here we have used rapid freezing techniques to avoid concerns about the possible effects of fixation on the *in vivo* form of myosin¹⁰ and demonstrate the existence of a myosin filament lattice in relaxed vascular smooth muscles in which we also show myosin light chain phosphorylation to be minimal (<5%).

Longitudinal strips of rabbit portal anterior mesenteric vein were attached to force transducers, stretched to physiological (1.75× slack) length, mounted in muscle baths at 37°C in oxygenated Krebs solution and equilibrated for ~1 h. Some muscles were exposed for 1.5 min to 0.2 $\mu\text{g ml}^{-1}$ isoprenaline which is known to hyperpolarize and relax smooth muscle¹¹. The muscle bath was quickly lowered and a beaker of supercooled Freon 22 (ref. 12) was shot up to the muscle. The moment of impact was seen on the tension record that showed that the

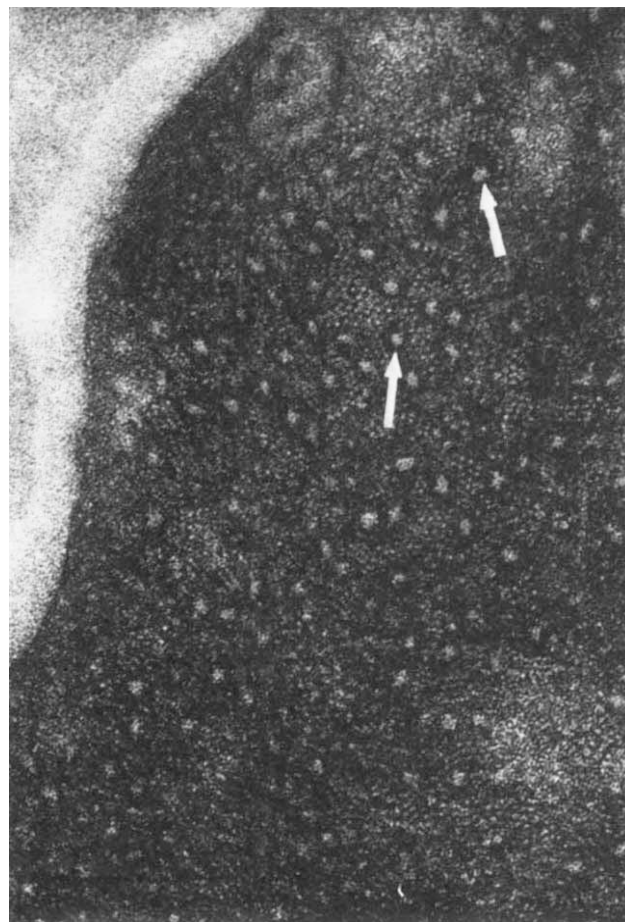


Fig. 1 Transverse section of a portion of an isoprenaline-relaxed PAMV smooth muscle cell, rapidly frozen and freeze substituted in acetone for 3 days at -80°C . Osmium crystals were added on warming and *en bloc* staining was done with uranyl acetate in methanol, which resulted in a negative staining effect in some of the outer cells. Arrows indicate myosin filaments surrounded by actin (thin) filaments. $\times 138,000$.

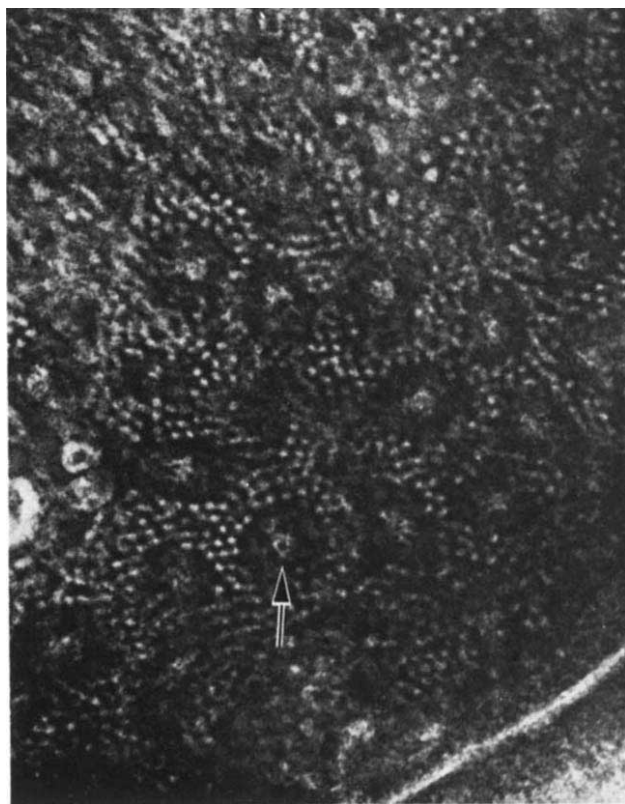


Fig. 2 Transverse cryosection of a rapidly frozen, briefly (15 min) glutaraldehyde-fixed muscle. The section was picked up on a drop of frozen sucrose in the cryo chamber, melted in a dry atmosphere and negatively stained with 1% ammonium molybdate. Myosin filaments (arrow) are surrounded by rosettes of actin filaments. $\times 200,000$.

muscle remained relaxed. The strips were processed by freeze substitution^{13,14} for ultrastructural studies or analysed for the degree of myosin light chain phosphorylation. The phosphorylated and unphosphorylated forms of the 20,000 molecular weight chain (LC₂₀) were separated by isoelectric focusing¹⁵ in urea-polyacrylamide gels followed by electrophoresis in SDS-containing slab gels in the second dimension¹⁶⁻¹⁸.

In relaxed, freeze-substituted muscles (Fig. 1), myosin filaments were consistently observed in the well frozen surface cells (nine muscles). The number and distribution of myosin filaments were similar in rapidly frozen and chemically fixed muscles^{3,19} (Figs 2, 3). The demonstration of thick filaments in the rapidly frozen freeze-substituted muscles renders highly unlikely the implication¹⁰ that the use of glutaraldehyde fixation before freezing (Fig. 2) induces the formation of an array of myosin filaments.

Regular arrays of actin filaments were more difficult to maintain in the rapidly frozen muscles, due to disorder caused by minute ice crystals. Therefore, three muscles were preincubated in a cryoprotectant—10% bovine serum albumin, 10% ethylene glycol or 10% dimethyl sulphoxide in Krebs solution—for 5 min before the addition of isoprenaline. Ethylene glycol and dimethyl sulphoxide abolished the spontaneous contractions. Ethylene glycol provided the greatest depth of cryoprotection and preserved the order of the actin filaments (Fig. 4). In these preparations, the ratio of actin to myosin filaments was 12:1, similar to the ratios found in conventionally fixed muscles^{3,19}.

The average LC₂₀ phosphorylation in portal veins relaxed with isoprenaline (six animals) was less than 5% (Fig. 5a). An

additional three relaxed portal veins which were cryoprotected before freezing (Fig. 4) showed similarly low levels of LC₂₀ phosphorylation as untreated resting muscles. In contrast to the relaxed muscles, significant levels of LC₂₀ phosphorylation, ranging over 23–73%, were measured in muscles frozen 2–15 s after stimulation with a high K⁺ depolarizing solution (Fig. 5b, c), with an average time-to-peak tension after activation of 12 s. The absence of any significant phosphatase activity during extraction of contractile proteins was demonstrated in control experiments using purified gizzard actomyosin in which LC₂₀ was ~80% phosphorylated (as shown by ³²P incorporation): phosphorylated actomyosin was (1) ground up with frozen taenia coli before extraction, (2) extracted separately or (3) ground up with frozen taenia coli and incubated at room temperature for 20 min before extraction. In (1) and (2) at most 4% of the ³²P label was released during the normal extraction procedure, while in case (3) almost all the labelled phosphate was released, as also demonstrated by the absence of phosphorylated LC₂₀ in the two-dimensional gels. Other control experiments showed that there was no change in LC₂₀ phosphorylation when the frozen homogenates were incubated for up to 4 h at room temperature in the solution (9.0 M urea, 5% mercaptoethanol) used for isoelectric focusing. These controls and the clear demonstration of phosphorylation of light chains in contracted muscle indicate that the extraction procedure and isoelectric focusing followed by SDS-gel electrophoresis provide an accurate estimate of the degree of LC₂₀ phosphorylation.

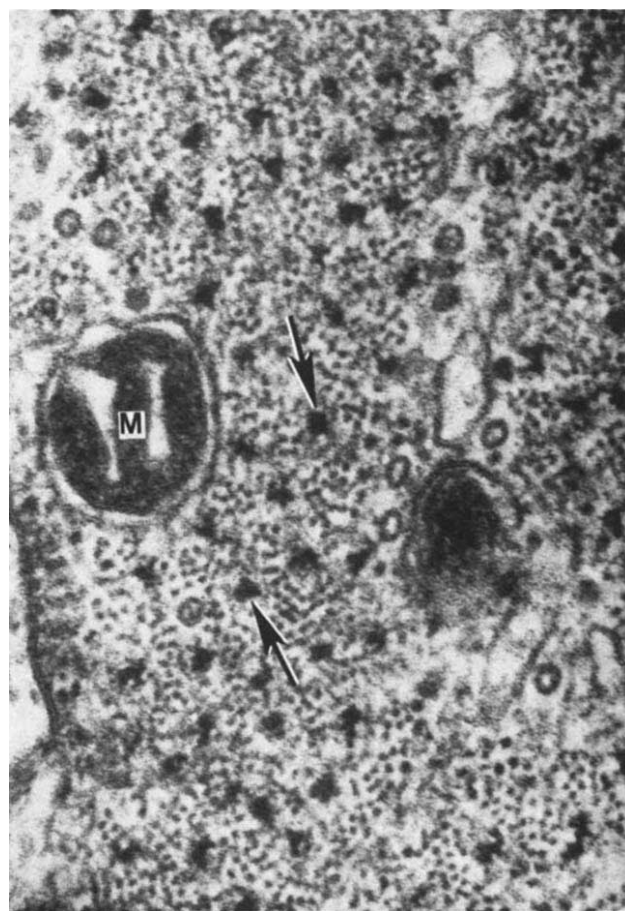


Fig. 3 Transverse section of an isoprenaline-relaxed smooth muscle, fixed in 2% glutaraldehyde followed by 2% tannic acid before post-fixation in osmium and *en bloc* staining with uranyl acetate. Arrows indicate myosin filaments; M, mitochondrion. $\times 148,000$.

That most of the myosin in mature smooth muscle is in filamentous form is also supported by the close agreement between estimates of myosin content obtained by SDS gels^{7,20-22} and by quantitative electron microscopy^{3,19}. The same actin:myosin ratios in rabbit portal anterior mesenteric vein were obtained when measured by either SDS-gel electrophoresis or quantitative electron microscopy (P. F. Berner, A. V. S., A. P. S. and H. Holtzer, unpublished observations). We also note that short filaments ($\sim 0.5 \mu\text{m}$) are produced by phosphorylation of myosin *in vitro*, unlike the $2.2\text{-}\mu\text{m}$ long filaments found in smooth muscle *in situ*¹⁹.

Thus, there is compelling evidence that myosin in resting smooth muscle is predominantly filamentous, even when the light chains are unphosphorylated, although we do not wish to extrapolate our findings to non-muscle cells. The rapid freezing techniques (refs 12, 23-25 and present study) that circumvent the objections that chemical fixation and associated changes in ionic strength, cation and nucleotide concentrations may induce the formation of myosin filaments also seem to be the most direct method of exploring the *in situ* state of myosin in non-muscle cells.

We thank Dr S. Chacko for providing ^{32}P -labelled turkey gizzard actomyosin. This work was supported by NIH grant HL15835 to the Pennsylvania Muscle Institute.

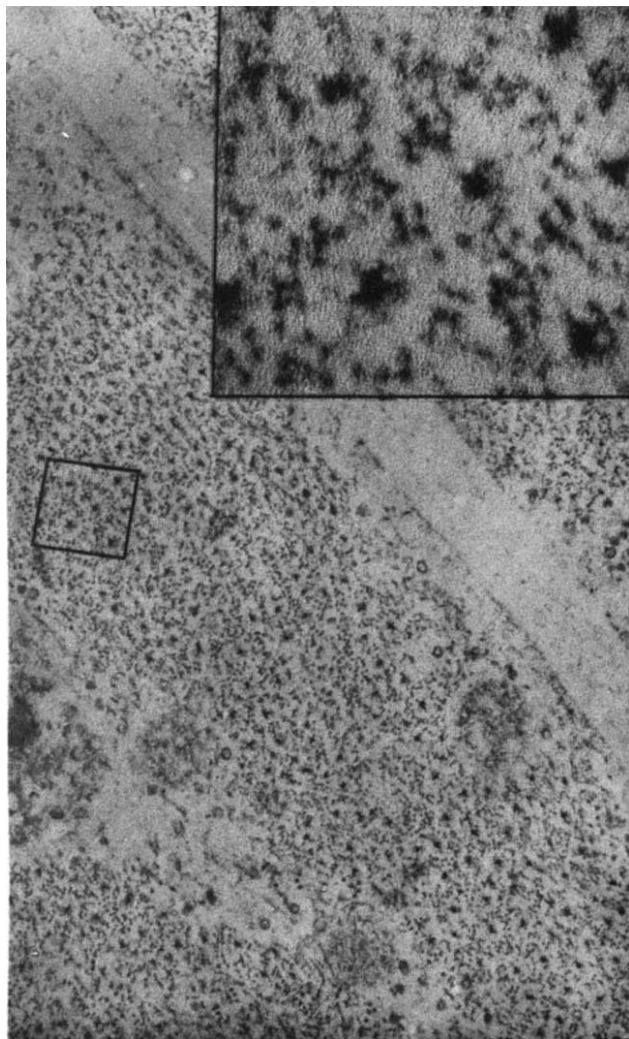


Fig. 4 Transverse section of a rapidly frozen isoprenaline-relaxed smooth muscle which has been cryoprotected with 10% ethylene glycol. $\times 60,000$. Inset: regular arrays of thick filaments occur which are surrounded by actin filaments shown at higher magnification ($\times 330,000$). LC_{20} phosphorylation in this muscle was $< 5\%$.

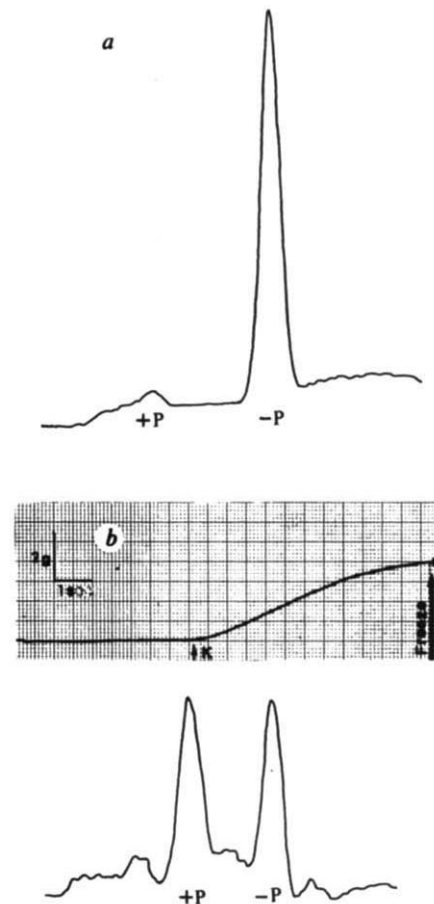


Fig. 5 *a*, Densitometer trace of a gel of an isoprenaline-relaxed muscle showing less than 5% phosphorylation of the LC_{20} . *b*, Tension trace of a portal vein exposed to a high K^+ depolarizing solution. The muscle was rapidly frozen at 6 s (arrow) by shooting a beaker of supercooled Freon 22 up to the muscle. *c*, Densitometer trace of the LC_{20} myosin light chains in an aliquot of the same, contracted muscle, showing that 50% of the LC_{20} s are in the phosphorylated form. For the phosphorylation studies, the frozen muscles were pulverized with 0.5 M HClO_4 at liquid N_2 temperature. After thawing, the protein precipitate was dissolved in 9 M urea and 5% mercaptoethanol.

Received 21 July; accepted 8 October 1981.

- Kelly, A. M. & Rice, R. V. *J. Cell Biol.* **42**, 683-694 (1969).
- Shoenberg, C. F. *Tissue Cell* **1**, 83-96 (1969).
- Devine, C. E. & Somlyo, A. P. *J. Cell Biol.* **49**, 636-649 (1971).
- Somlyo, A. V. in *Handbook of Physiology. Vascular Smooth Muscle* (eds Bohr, D. F., Somlyo, A. P. & Sparks, H. V.) 33-67 (American Physiological Society, Williams & Wilkins, Baltimore, 1980).
- Campbell, A. R. & Chamley, J. H. *Cell Tissue Res.* **156**, 201-216 (1975).
- Gabella, G. in *Smooth Muscle* (eds Bulbring, E., Brading, A. F., Jones, A. W. & Tomita, T.) 1-46 (Edward Arnold, London, 1981).
- Small, V. J. & Sobieszek, A. *Int. Rev. Cytol.* **64**, 241-306 (1980).
- Suzuki, H., Onishi, H., Takahashi, K. & Watanabe, S. *J. Biochem.* **84**, 1529-1542 (1978).
- Scholey, J. M., Taylor, K. A. & Kendrick-Jones, J. *Nature* **287**, 233-235 (1980).
- Shoenberg, C. F. & Stewart, M. J. *Muscle Res. Cell Motility* **1**, 117-126 (1980).
- Somlyo, A. V., Haeusler, G. & Somlyo, A. P. *Science* **169**, 490-491 (1970).
- Somlyo, A. V., Shuman, H. & Somlyo, A. P. *J. Cell Biol.* **74**, 828-857 (1977).
- Bullivant, S. *Lab. Invest.* **14**, 440-457 (1965).
- Franzini-Armstrong, C., Heuser, J. E., Reese, T. S., Somlyo, A. P. & Somlyo, A. V. *J. Physiol., Lond.* **283**, 133-140 (1978).
- DiSalvo, J., Gruenstein, E. & Silver, P. *Proc. Soc. exp. Biol. Med.* **158**, 410-414 (1978).
- O'Farrell, P. H. *J. Biol. Chem.* **250**, 4007-4021 (1975).
- Driska, S. P., Aksoy, M. O. & Murphy, R. A. *Am. J. Physiol.* **240**, C222-C233 (1981).
- Butler, T. M. & Siegman, M. J. *Fedn Proc.* (in the press).
- Ashton, F. T., Somlyo, A. V. & Somlyo, A. P. *J. molec. Biol.* **98**, 17-29 (1975).
- Tregear, R. T. & Squire, J. M. *J. molec. Biol.* **77**, 279-290 (1973).
- Murphy, R. A., Herlihy, J. T. & Megerman, J. *J. gen. Physiol.* **64**, 691-705 (1974).
- Cohen, D. M. & Murphy, R. A. *J. gen. Physiol.* **72**, 369-380 (1978).
- Somlyo, A. V. & Silcox, J. in *Microbeam Analysis in Biology* (eds Lechene, C. & Warner, R.) 535-555 (Academic, New York, 1979).
- Heuser, J. E., Reese, R. S. & Landis, D. M. D. *Cold Spring Harb. Symp. quant. Biol.* **40**, 17-24 (1976).
- Sjöström, M. in *Microbeam Analysis in Biology* (eds Lechene, C. & Warner, R.) 427-447 (Academic, New York, 1979).

Stress does not alter the conformation of a domain of the myosin cross-bridge in rigor muscle fibres

Roger Cooke

Department of Biochemistry and Biophysics and the Cardiovascular Research Institute, University of California, San Francisco, California 94143, USA

The force of muscle contraction is thought to be generated by a change in the effective angle of a myosin cross-bridge while it is attached to an actin filament. Thus studies of the relationship between this conformation and force provide insight into the mechanism of contraction. Recently it has been shown¹ that paramagnetic probes can be attached selectively and rigidly to a reactive sulphhydryl on the myosin cross-bridge and that their angular distribution can be measured with great accuracy. In rigor skeletal fibres the probes are highly ordered relative to the fibre axis. Here I show that forces of up to 0.2 N mm^{-2} cause no change in the angular distribution of the probes. Thus there is a domain of the cross-bridge whose conformation is not influenced by stress, and this result places restrictions on the location of elastic and force-generating elements within the cross-bridge.

The heads of the myosin molecule form bridges between the myosin and actin filaments. During contraction several changes in the conformation of these cross-bridges are thought to occur, including the stretching of an elastic element and a change in the configuration of some force-generating element resulting in a large ($\sim 45^\circ$) change in the effective angle of the cross-bridge^{2,3}. The location of these elements within the cross-bridge is unknown; one possibility is that one or both may reside in the actin-myosin bond. By measuring the angle of one portion of myosin in a rigor fibre, that is, a fibre in the absence of ATP, we provide information on the location of these elements.

Various methods have been used to monitor the angle of the myosin heads in muscle fibres, including electron microscopy⁴, X-ray diffraction⁵ and fluorescence spectroscopy⁶. Only two of these have been used to explore the relationship between force and cross-bridge orientation in rigor fibres. Naylor and Podolsky⁷ have shown that the intensity ratio of two equatorial reflections in the X-ray diffraction pattern did not change on application of a static strain, and dos Remedios *et al.*⁸ showed that the polarization of tryptophan fluorescence did not change when the rigor fibre was subjected to a periodic strain. For both these methods, the parameters observed reflected the conformation of the entire fibre, the X-ray being sensitive to the distribution of mass in the filament array and the fluorescence being derived from all tryptophans in the fibre. Thus, paramagnetic probes, which allow accurate and direct measurement of the angular conformation of one portion of the myosin molecule, provide quite different information from that obtained using these other methods.

Two probes were used having reactive groups that resembled iodoacetamide (IASL) and melanimide (MSL). Both labels have been shown to react selectively with a sulphhydryl on the myosin head, designated SH₁ (refs 2, 9). The hyperfine interaction between the unpaired electron and the nitrogen nucleus splits the spectrum of a nitroxide radical into three lines. Both the splitting between these lines and the g value depend on the angles between spin probe axes and the magnetic field of the spectrometer^{2,10}. When the magnetic field is aligned along the fibre axis this effect allows measurement of the angle between this axis and the principle axis of the spin probe.

Figure 1 shows the spectra of fibres labelled with IASL. The fibres are in rigor at a sarcomere length of $2.2\text{--}2.4 \mu\text{m}$ where all the myosin heads are overlapped by thin filaments. These spectra are seen to consist essentially of three lines, indicating that most of the probes have the same approximate angle with respect to the fibre axis. It was shown previously² that such spectra are well described by assuming that the angular distribution of probes is gaussian, centred at 68° with a full width at half height of 15° . As shown in Fig. 1, a static stress applied to the fibres caused little alteration in their spectra. It was expected that stress would result in: (1) a shift in the average angle of the gaussian distribution; (2) a broader distribution and/or (3) that the cross-bridges would assume some new and radically different distribution. Any shift in the mean angle of the gaussian distribution would result in a shift in the splitting between the peaks. This splitting, indicated by arrows in Fig. 1, is shown in Fig. 2 for both labels as a function of the stress on the fibre. There was no change in the mean angle of either probe within

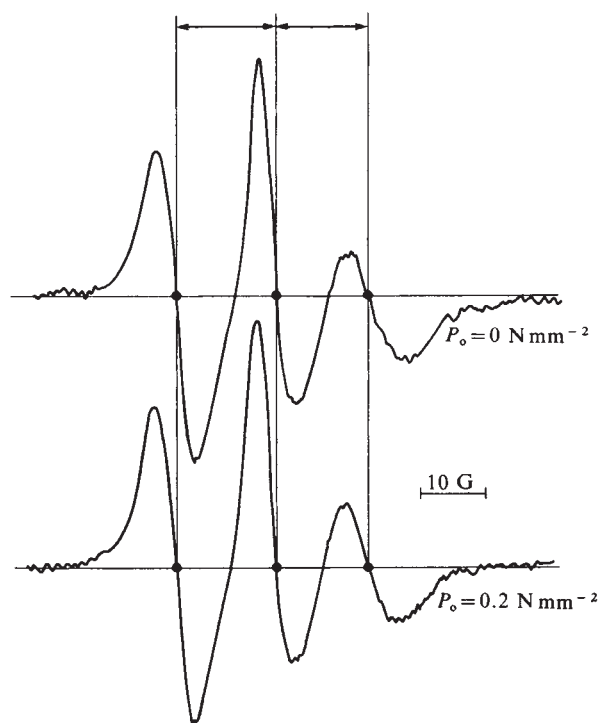


Fig. 1 Paramagnetic spectra of IASL attached to rigor fibres shown for two cases: above, tension = 0; below, tension = 0.2 N mm^{-2} . The fibres were labelled with 0.5 mM IASL for 1 h following procedures described elsewhere¹. Several methods, including quantitation of peak heights in the spectrum and the decrease in the EDTA-ATPase of myosin extracted from the fibres, indicated that 70–100% of the myosin heads were labelled specifically on the reactive SH by this procedure. The rigor stiffness and the ability to generate isometric tension were both unchanged by labelling of the myosin heads. Spectra were obtained using a Varian E-3 spectrometer with a cavity modified to permit simultaneous recording of tension and spectra. This was accomplished by drilling holes (2 mm diameter) in the side faces of the cavity. A fibre bundle (6–10 fibres) was mounted in a capillary that was inserted through the cavity, so that the long axis of the bundle was aligned along the magnetic field of the spectrometer. One end of the fibre was secured by surgical silk to a solid-state tension transducer mounted on one side of the cavity. A force was exerted on the other end of the fibre using weights that were attached by a length of surgical silk. Each spectrum was acquired in 4 min, and during this time the fibre did not lengthen appreciably. The derivative of the absorption is plotted as a function of the magnetic field strength, thus the splitting between the three peaks is given by the distance between points where the spectrum crosses the base line, indicated by arrows above the spectra.

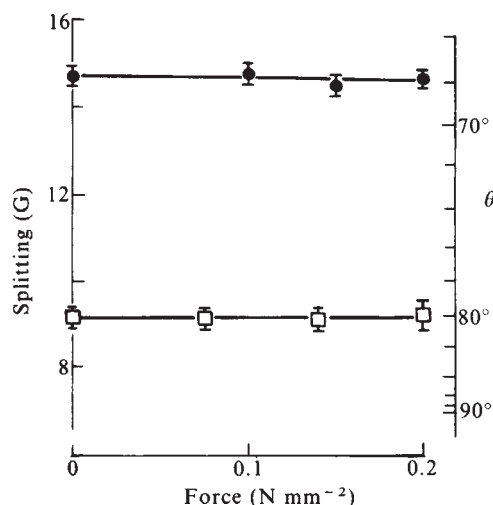


Fig. 2 The splitting between the lines in the electron paramagnetic resonance spectra, indicated by arrows in Fig. 1, is plotted as a function of the stress applied to the fibre. This splitting is a function of θ , the mean angle of the angular distribution of the probes, shown on the right. ●, Data from fibres labelled with IASL; □, data from fibres labelled with MSL. Fibres were labelled with MSL using procedures outlined previously¹ except that the concentration of $K_3Fe(CN)_6$, used to reduce nonspecific labels, was reduced to 25 mM. Each data point represents the mean of at least two separate measurements for each of two to four different fibres. Because of errors in the measurement of the cross-section of the fibre bundles, there is a 20% error in the tension measurements.

the experimental accuracy of the method ($\pm 1^\circ$) (Fig. 2). The width of the lines in Fig. 1 is related to the spread of the angular distribution. The experimental apparatus used here gave a sharp angular distribution with a full width at half height of $\sim 12^\circ$. The application of a stress to the fibres resulted in a slight sharpening of the lines, equivalent to further reducing the width of the gaussian distribution by $\sim 2^\circ$. This effect was found after a small stress was applied, and thus is probably due to better alignment of the fibres than in previous studies². If some of the probes had changed to a new angle, additional intensity would be generated at new positions in the spectrum. No new intensity was seen and a careful comparison of spectra indicated that $<10\%$ of the probes changed angles from the gaussian distribution characteristic of rigor muscle.

Although the paramagnetic spectra allow the accurate measurement of the angle of the probe relative to the fibre axis, the relative orientation of the probe and the myosin head is not known with certainty. This uncertainty prompts the question, could the head rotate without probe rotation? Some information on the orientation of the probe relative to the head can be obtained from the measurement of the rotational relaxation time of probes bound to myosin subfragment 1 free in solution. These probes have a long relaxation time of ~ 200 ns, indicating that their axis lies along some long axis of subfragment 1 (ref. 11). As the myosin head is thought to be an elongated structure with an actin binding site at one end and the attachment to the thick filament at the other, any probe whose axis lies along the long axis of the head should be sensitive to rotations of the head induced by stress. In addition, two paramagnetic probes, linked to the SH_1 through structurally different groups and having axes at least 12° apart, gave the same result. Thus we conclude that the region of myosin surrounding the SH_1 does not change orientation when a stress is applied to a rigor fibre. These results agree with those of others^{7,8} who also concluded that the myosin head did not rotate on application of stress.

When a step change in length is applied to an isometrically contracting fibre, there is a complex transient in tension of the

fibre^{3,12}. The first portion of this transient, a sudden change in tension, is interpreted as arising from an elastic element that is in series with the element which generates force. Recent measurements of these transients at different sarcomere lengths have shown that this elasticity resides in the cross-bridge structure and thus may constitute an important part of the contractile apparatus¹². Some theorists have placed this elastic element in the bond between actin and myosin¹¹, in which case the application of a static stress would change the angle of the cross-bridge. Thus our results show that this element cannot reside in the acto-myosin bond and it must lie somewhere between the SH_1 probe site and the thick filament. As the elastic element is a part of the cross-bridge it may be located in the S2 region which connects the head to the filament backbone, or it may result from some flexibility in the head.

Present models of the contractile interaction envisage that force is produced by a myosin head attached to actin when the head rotates through an axial gradient of free energy. Whether the contractile element will shift its configuration on application of stress to rigor muscle depends on the shape of the free energy curve for the state in which no nucleotides are bound to myosin. If the configuration of the contractile element is confined in a deep well of free energy in this state then no change is expected. However, if it is confined by a shallower well, such as proposed by some theorists for other states¹³, it should change configuration. Therefore my results require that, unless the contractile element is very stiff in the rigor state, it must also lie between the probe site and the thick filament. This suggests that the actin-myosin bond is stiff and that it is not involved directly in the changes in cross-bridge conformation discussed above.

I conclude that there is a domain of the myosin head whose conformation is not influenced by force on the rigor fibre. As the SH_1 is in this domain, its position provides some estimate of the size of the domain. Although the exact location of the SH_1 is unknown, fluorescent energy transfer has provided some information on its relation to the actin binding site. The distance from SH_1 to Cys 373 of actin has been found to be ~ 6 nm (ref. 14). The cys 373 is 3.5 nm from the centre of the actin filament whose diameter is estimated to be 7–8 nm (ref. 15), and thus it cannot be very far from the actin-myosin binding site, also shown to be on the outer portion of the actin filament. If these distances are laid out on a three-dimensional model of the structure of the acto-S1 complex¹⁶, it appears unlikely that the SH_1 could be directly adjacent to the actin binding site, and thus the domain discussed above probably comprises a significant portion of the myosin head.

I thank Aron Yoffe and Laurence McCabe for assistance in completing the experiments, and Drs David Thomas, Michael Wilson, Robert Mendelson and Manuel Morales for their helpful discussions. This work was supported in part by grants from the USPHS, HL16683 and Research Career Development Award AM00479.

Received 22 June; accepted 14 October 1981.

1. Thomas, D. D. & Cooke, R. *Biophys. J.* **32**, 891 (1980).
2. Cooke, R. & Franks, K. *Biochemistry* **19**, 2265 (1980).
3. Huxley, A. F. & Simmons, R. M. *Nature* **233**, 533 (1971).
4. Reedy, M. K., Holmes, K. C. & Tregear, R. T. *Nature* **207**, 1276 (1965).
5. Huxley, H. E. & Brown, W. J. *molec. Biol.* **30**, 383 (1967).
6. Tregear, R. T. Mendelson, R. A. *Biophys. J.* **15**, 455 (1975).
7. Naylor, G. R. S. & Podolsky, R. J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
8. dos Remedios, C. G., Millikan, R. C. G. & Morales, M. F. *J. gen. Physiol.* **59**, 103 (1972).
9. Seidel, J. C., Chap, M. & Gergely, J. *Biochem. J.* **9**, 3265 (1970).
10. McCalley, R. C., Shimshick, E. J. & McConnell, H. M. *Chem. Phys. Lett.* **13**, 115 (1972).
11. Thomas, D. D., Seidel, J. C., Hyde, J. S. & Gergely, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1729 (1975).
12. Ford, L. E., Huxley, A. F. & Simmons, R. M. *J. Physiol., Lond.* **311**, 219 (1981).
13. Eisenberg, E., Hill, T. L. & Chen, Y. D. *Biophys. J.* **29**, 195 (1980).
14. Takashi, R. *Biochem. J.* **18**, 5164 (1979).
15. Taylor, D. L., Reidler, J., Spudich, J. A. & Stryer, L. *J. Cell Biol.* **89**, 362 (1981).
16. Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 17a (1970).

Synthesis of an unspliced cytoplasmic message by an adenovirus 5 deletion mutant

Leon Carlock & Nicholas C. Jones*

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

Most genes in eukaryotic cells and their viruses contain non-coding intervening sequences (introns)¹⁻⁵. These sequences are excised from each transcript during maturation of mRNA by a process known as RNA splicing. Although very little is known about the mechanism by which splicing occurs or its biological significance, several studies have suggested that the removal of introns is a necessary prerequisite to the accumulation of stable, cytoplasmic mRNA⁶⁻¹⁰. Here we describe a mutant of adenovirus 5 (Ad5) containing a deletion within early region Ela; the deletion prevents splicing of the normal Ela transcripts. The unspliced transcript is found in the cytoplasm of infected cells, showing that for this gene, splicing is not necessary for the biosynthesis of cytoplasmic message.

Region Ela is located at the extreme left-hand end of the adenovirus genome (1.5-4.5 map units). At early times of infection two overlapping mRNAs (a 12S and a 13S species) are produced by differential splicing of a common precursor^{11,12}. The two messages share 5' and 3' sequences but differ in the extent to which internal sequences are removed by RNA splicing (Fig. 1). The structure of the *dl502* genome is also shown in Fig. 1. DNA sequence analysis identified a deletion of 330 base pairs fusing nucleotide 1,009 to nucleotide 1,340. The deletion thus removes the acceptor and one of the donor (D2) splice junctions together with flanking coding sequences.

Because the common acceptor splice junction is absent in *dl502*, it was anticipated that the primary transcript from this region would fail to undergo splicing. This was found to be the case. Cytoplasmic RNA isolated from wild-type and *dl502*-infected HeLa cells was analysed by the S₁ mapping procedure of Berk and Sharp¹³ and as the results described in Fig. 2 clearly show, an unspliced Ela message is synthesized and accumulates in the cytoplasm of *dl502*-infected cells. As the hybridizations are performed with an excess of DNA probe, the intensity of autoradiograph bands is proportional to the abundance of the corresponding mRNA. Comparison of the Ela-generated bands from wild-type and *dl502* RNA preparations indicates that the unspliced Ela product from *dl502* accumulates in the cytoplasm to 25-50% of the levels of wild-type Ela products. The lower levels are probably due to a defective Ela gene product in *dl502*. Other mutants with altered Ela products show a similar decrease in the levels of Ela messages, possibly indicating that an Ela product regulates its own synthesis^{14,15}.

We consider it very unlikely that the unspliced message detected could have resulted from nuclear leakage during fractionation of the RNA. The RNA was isolated using isotonic

RNA extraction conditions, where nuclear leakage is reported to be low^{16,17}. Furthermore, two phenotypic properties of *dl502* that we have studied indicated that a defective Ela gene product is being synthesized and is partially functional. The first property is the host-range phenotype of *dl502*. Not unexpectedly, *dl502* is defective for growth on HeLa cells. Region Ela encodes a product required for productive growth^{18,19}; mutants with alterations in this region, including *dl502*, are grown on 293 cells. These are Ad5-transformed human embryonic kidney cells²⁰; they contain and express the left-hand end 11% of the Ad5 genome and consequently will complement mutants which have alterations in this region. *dl502* has a plaquing ratio on HeLa versus 293 cells of $\sim 10^{-4}$; this is 10^2 - 10^3 -fold higher than the ratio obtained with another mutant, *dl312*, which contains a deletion removing almost the entire Ela region (data not shown and ref. 18). We attribute this difference to the synthesis of a partially functional Ela gene product by *dl502*, supporting our conclusion that a message from this region does accumulate in the cytoplasm of *dl502*-infected cells.

The second phenotypic property suggesting that a partially functional Ela product is being synthesized by *dl502*, is that transcripts from other regions of the genome are detected, although at reduced levels. A region Ela gene product is known to regulate the expression of other early regions of the genome, regions E1b, E2, E3 and E4^{21,22}. Transcripts from these regions are not detected in *dl312*-infected cells unless very high MOIs are used¹⁵. However, as shown in Figs 2 (for region E1b) and 3 (for regions E2, E3 and E4), transcripts from all four of these regions are detected in the cytoplasm of HeLa cells infected with *dl502* at relatively low MOIs (~ 20 plaque-forming units (PFU) per cell). We again attribute this difference between *dl502* and *dl312* to the synthesis of a partially functional Ela gene product by *dl502*.

The major finding of the present study is that an unspliced region Ela transcript accumulates in the cytoplasm of *dl502*-infected cells, showing that splicing and the removal of intron sequences are not necessary prerequisites to the production of stable, cytoplasmic RNA from this particular gene. The characterization of SV40 mutants containing deletions in both early and late genes has led to the conclusion that splicing is required not only for the removal of noncoding intron sequences but also in some unknown way for the stabilization and/or transport of message⁶⁻⁸. However, it is now known that several genes, including the gene for adenovirus protein IX²³ and a human interferon²⁴, do not contain introns; the synthesis of message from these genes does not therefore require splicing. In addition, two other cases recently have been described where RNA splicing, although normally occurring during the synthesis of message from the genes examined, was found not to be obligatory to the synthesis of message from those genes^{25,26}. A viable mutant of SV40, containing a small deletion in the leader region of the late genes, accumulated an unspliced 19S mRNA in the cytoplasm of infected cells²⁵. Interestingly, other mutants containing alterations in different locations and which also prevented splicing of the 19S precursor RNA did not accumulate 19S message⁶, suggesting that the location of the deletion was critical in determining whether splicing was obligatory for

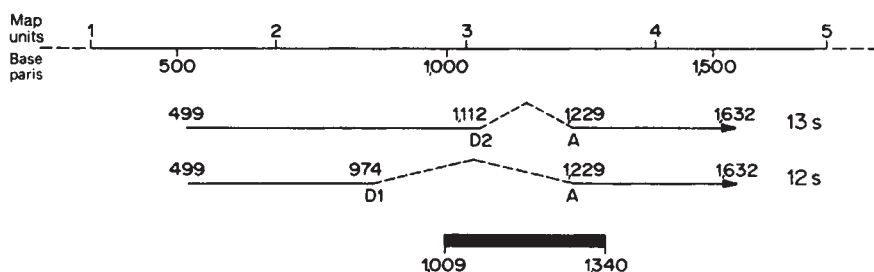


Fig. 1 Structure of the *dl502* genome. The position of the deletion in *dl502* is shown with respect to the two mRNA species produced from this region early after infection with Ad5^{11,12}. The solid line indicates sequences present in the mRNA, caret symbols represent intervening sequences removed by RNA splicing and the arrows mark the position of the 3' end of the RNAs. The nucleotide positions of the splice junctions and 5' and 3' termini are from Perricaudet *et al.*²⁷. D1 and D2 are the terms given to the two donor splice junctions and A is the acceptor junction. The deletion in *dl502* as determined by DNA sequence analysis is indicated by the solid bar.

Fig. 2 S_1 analysis of cytoplasmic RNA from wild-type and *dl502*-infected cells. HeLa cells were infected with purified virions at a MOI of ~ 20 PFU per cell. Cytosine arabinoside was added 1 h later and the cells were collected 8 h post-infection. Cytoplasmic RNA was isolated from the infected cells by the isotonic extraction procedure described previously²¹. The RNAs were analysed by the S_1 mapping procedure essentially as described by Berk and Sharp¹³. 32 P-labelled DNA fragments (5 μ g full genome equivalents per ml; specific activity $2\text{--}4 \times 10^6$ c.p.m. per μ g) were hybridized to total cytoplasmic RNA (5 mg ml⁻¹) in 80% formamide, 0.4 M NaCl, 0.04 M PIPES pH 6.4, 0.001 M EDTA at 59°C for 3 h. The hybridization mixture was then digested with S_1 endonuclease (40 units) in 200 μ l, 0.03 M NaOAc, 0.25 M NaCl, 0.001 M ZnCl₂, 5% glycerol at 37°C for 30 min. The digestion products were ethanol precipitated and electrophoresed on 3.8% polyacrylamide gels containing 8 M urea. *a*, RNA from wild-type infected cells hybridized to the *Bgl*II-D fragment (0–9 map units) isolated from wild-type DNA. Four major S_1 -resistant bands are detected. The band 1,630 nucleotides in length results from hybridization of region E1b mRNA (a region extending over 4.6–11.2 map units¹²) to the probe; the other three bands result from hybridization of *E1a* mRNA. *b*, RNA from *dl502*-infected cells hybridized to the same probe as that used in lane *a*. A 1,630-nucleotide E1b band is seen together with two additional bands, 511 and 387 nucleotides long. These are the expected products if an unspliced RNA from region E1a was synthesized and hybridized to the probe; a deletion loop extending over 2.8–3.8 map units would be present in the hybrid and be digested with the S_1 endonuclease to give the two fragments seen. *c*, RNA from wild-type-infected cells hybridized to wild-type *Hind*III-G fragment (0–7.8 map units). The band resulting from the hybridization of E1b mRNA is smaller with this probe than the one used in lane *a* (1,108 nucleotides). *d*, RNA from *dl502*-infected cells hybridized to the *dl502* *Hind*III-G fragment. As well as the band derived from E1b mRNA, a single additional band of ~ 800 nucleotides in length was detected which again is the size of the probe sequence that would be protected by an unspliced E1a mRNA. A diagram of region E1 and the mRNAs synthesized from this region by wild-type virus and *dl502* is shown. The wild-type RNAs are described in detail in Fig. 1 legend.

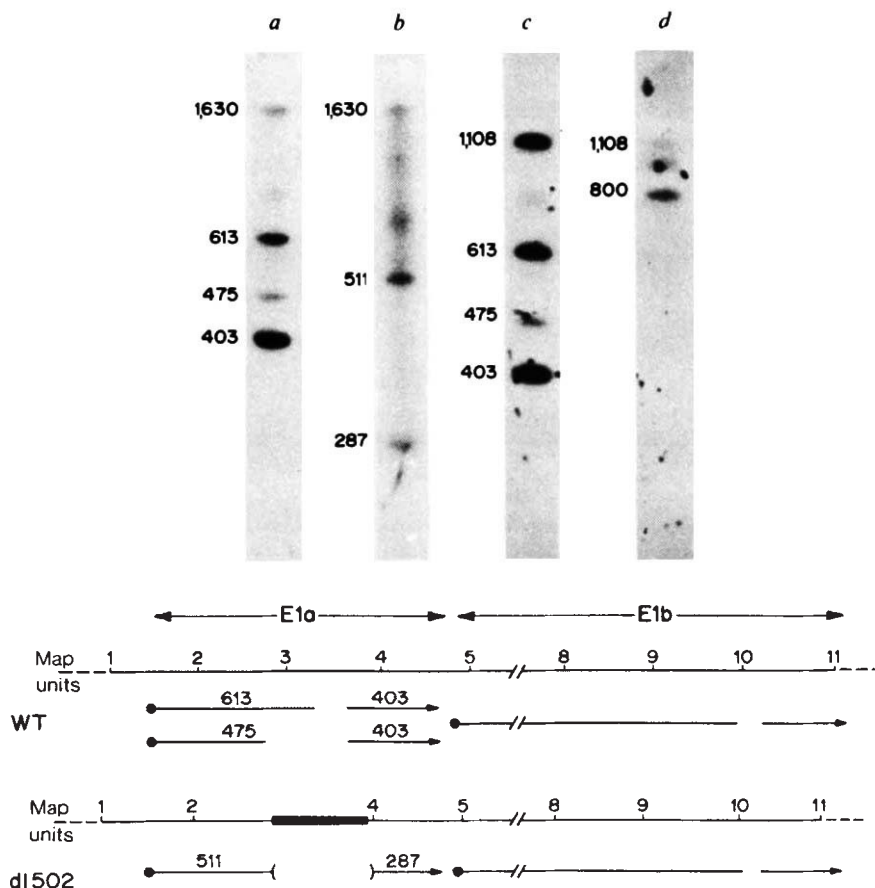
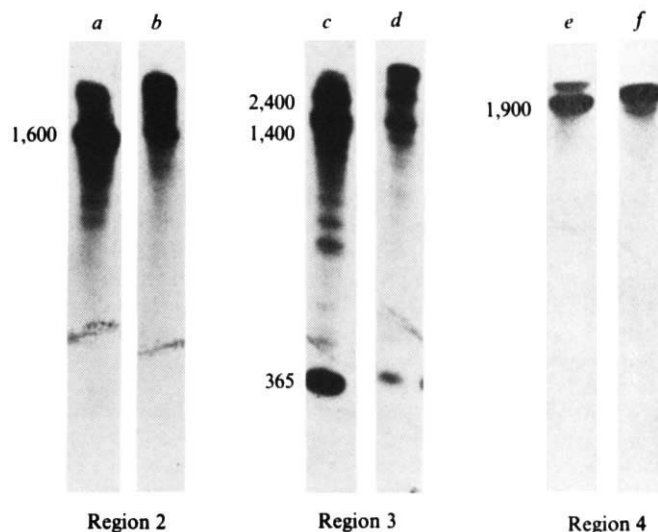


Fig. 3 S_1 analysis of cytoplasmic RNA from wild-type and *dl502* infected cells using probes specific for early regions 2, 3 and 4. The isolation of RNA and the S_1 mapping procedure are described in Fig. 2 legend. Lanes *a*, *b*, hybridization of RNA from wild-type (*a*) and *dl502* (*b*) infected cells to the *Hind*III-A fragment (50.1–72.8 map units). A major product 1,600 nucleotides long is normally found. Lanes *c*, *d*, hybridization of wild-type (*c*) and *dl502* (*d*) RNA to the *Hind*III-B fragment (72.8–89.1 map units). Three major products 2,400, 1,400 and 365 nucleotides in length are normally found. Lanes *e*, *f*, hybridization of wild-type (*e*) and *dl502* (*f*) RNA to the *Hind*III-F fragment (89.1–97.1 map units). A 1,900-nucleotide band is the major product. The largest bands detected in all the hybridizations are due to renaturation of the DNA probe; the renaturation is particularly evident in the hybridizations involving *dl502* RNA.



message production. It is therefore possible that it is not splicing *per se* that is the critical factor in producing stable message but the presence or absence of certain important sequences in the precursor.

One possible model to explain all the results would be that a sequence(s) in the precursor RNA is necessary for the synthesis of mature cytoplasmic message (for example by mediating an interaction between the RNA and the nuclear membrane where the transport apparatus and possibly the splicing apparatus resides); if this sequence is missing maturation of the message

from the precursor could not occur. In genes that do not contain introns such a sequence would be located within coding regions. In genes that do contain introns the sequence could be in the intron, the exons or partially in both (perhaps spanning a splice junction). In examples where deletions prevented the accumulation of messages, the model would predict that the necessary sequences had been altered or removed. In examples such as *dl502*, however, where unspliced messages are found, the prediction would be that the sequence remained and was located within the coding sequences or the remaining portion of the

intron sequences. Implicit in the model is that splicing *per se* is not necessarily linked to the synthesis of cytoplasmic RNA. Nevertheless, in the normal situation unspliced precursors are confined to the nucleus. This would be the case if the rate of splicing far exceeded that of transport. This model would also be feasible if a particular feature of the secondary structure of the

precursor RNA was important for message biosynthesis and not simply a primary RNA sequence. The isolation and characterization of additional mutants would help distinguish between these possibilities.

We thank John Williams for technical assistance. This work was supported by American Cancer Society grant NP-306.

Received 31 July; accepted 16 October 1981.

1. Abelson, J. A. *Rev. Biochem.* **48**, 1035-1069 (1979).
2. Darnell, J. E. *Prog. Nucleic Acid. Rec. molec. Biol.* **22**, 327-353 (1979).
3. Crick, F. H. C. *Science* **204**, 264-271 (1979).
4. Lewin, B. *Cell* **22**, 324-326 (1980).
5. Sharp, P. A. *Cell* **23**, 643-646 (1981).
6. Lai, C. J. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 71-75 (1979).
7. Khoury, G. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 41-54 (1979).
8. Gruss, P., Lai, C. J., Dahr, R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4317-4321 (1979).
9. Hamer, D. H. & Leder, P. *Cell* **18**, 1299-1302 (1979).
10. Wickens, M. P., Woo, S., O'Malley, B. W. & Gurdon, J. B. *Nature* **285**, 628-634 (1980).
11. Berk, A. J. & Sharp, P. A. *Cell* **14**, 695-711 (1978).
12. Chow, L. T., Broker, T. R. & Lewis, J. B. *J. molec. Biol.* **134**, 265-303 (1979).
13. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).

14. Berk, A. J., Lee, F., Harrison, T., Williams, J. & Sharp, P. A. *Cold Spring Harb. Symp. quant. Biol.* **44**, 429-436 (1979).
15. Ross, S. R., Levine, A. J., Galos, R. S., Williams, J. & Shenk, T. *Virology* **103**, 475-492 (1980).
16. Levis, R. & Penman, S. *Cell* **11**, 105-113 (1977).
17. Villarreal, L. P. & Jovanovich, S. J. *J. Virol.* **36**, 595-600 (1980).
18. Jones, N. & Shenk, T. *Cell* **17**, 683-689 (1979).
19. Harrison, T., Graham, F. & Williams, J. *Virology* **77**, 319-329.
20. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. J. *gen. Virol.* **36**, 59-72 (1977).
21. Jones, N. & Shenk, T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3665-3669 (1979).
22. Berk, A. J., Lee, F., Harrison, T., Williams, J. & Sharp, P. A. *Cell* **17**, 935-956 (1979).
23. Alestrom, P. *et al. Cell* **19**, 671-681 (1980).
24. Nagata, S., Mantei, N. & Weissman, C. *Nature* **287**, 401-408 (1980).
25. Gosh, P. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 1386-1390 (1981).
26. Treisman, R., Novak, V., Favaloro, J. & Kamen, R. *Nature* **292**, 595-600 (1981).
27. Perricaudet, M., Akusjarvi, G., Virtanen, A. & Pettersson, U. *Nature* **281**, 694-696 (1979).

Specific integration of REV proviruses in avian bursal lymphomas

M. Reza Noori-Daloii, Robert A. Swift & Hsing-Jien Kung*

Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824, USA

Lyman B. Crittenden & Richard L. Witter

US Department of Agriculture, Science and Education Administration, AR Regional Poultry Research Laboratory, East Lansing, Michigan 48824, USA

The most common naturally occurring cancer of chickens associated with retrovirus infection is lymphoid leukaemia (LL), a bursa (B) cell lymphoma. The primary causative agents are avian lymphoid leukaemia viruses (LLVs), which do not necessarily have an oncogene. Recent evidence of Hayward *et al.*¹ suggests that LLV infection promotes the expression of a cellular gene, *c-myc* (homologous to the oncogene of acute leukaemia virus, MC-29) thereby triggering the transformation process. In the majority of the tumours induced by LLV, the provirus is integrated next to the *c-myc* gene (refs 1, 3, 18, 19). To examine further the specific involvement of *c-myc* in lymphocytic transformation, we exploited our previous findings that replication-competent or non-defective reticuloendotheliosis virus (*nd* REV), genetically unrelated to LLV, are also capable of inducing lymphoma in chickens, with similar latent period and pathology to LL⁵. We have characterized the structure of the *nd* REV proviruses in the induced tumours and report here that proviral DNA is integrated next to *c-myc* in over 90% of the tumours analysed. This finding strengthens the hypothesis that the *c-myc* and its adjacent sequences are important in B-lymphocyte transformation. We also obtained evidence that amplification and structural alteration of the chromosomal region encompassing the REV provirus and *c-myc* gene have occurred during tumorigenesis in some of the tumours.

The DNAs extracted from bursal tumours induced by chick syncytial virus (CSV) strain of *nd* REV were digested with restriction endonucleases, electrophoresed in agarose gels and hybridized to probes specific for the *nd* REV genome (REV-probe) and the oncogene of MC-29 virus (MC-probe), respectively. The principal enzyme used in this study is *Eco*RI, which does not cleave the CSV DNA. As shown in Fig. 1C, lanes *a* and *b*, the unintegrated linear CSV DNA isolated from the acutely

infected cells displays the same migration pattern before and after *Eco*RI digestion. Similarly, the unintegrated circular DNA can also be shown to lack *Eco*RI sites (data not shown). When *Eco*RI was used to cleave the high molecular weight DNA from the tissues (Fig. 1A), in all samples including the uninfected control (lane U), a band of about 18 kilobases (kb) is detected which presumably represents the endogenous REV sequence (*endo*_{REV}). In the tumour sample (lanes 1-6), additional bands (indicated by dots) of differing sizes appear; these are referred to as tumour-specific (TS) bands. The TS band in lane 6, which has a mobility similar to that of *endo*_{REV}, is difficult to discern in this experiment. The endogenous sequence, however, is only weakly hybridizable to the REV-probe (presumably due to the divergence of these two sequences) and it is possible to differentiate the endogenous and exogenous REV sequences by conducting the probe hybridization in more stringent conditions. As illustrated in Fig. 1D, in the more stringent conditions (45 °C) the hybridization to *endo*_{REV} is reduced to the background level (see especially the uninfected control in lane U), whereas the intensity of the TS bands, including the 18-kb band of lane 6, are not affected. This analysis confirms the exogenous origin of all the TS bands.

To assess the linkage between REV proviruses and the *c-myc* gene, the *Eco*RI cleaved bursal DNAs were hybridized with the MC probe. As shown in Fig. 1B, the TS bands (indicated by dots) detected by the MC probe match the corresponding fragments identified by the REV probe in Fig. 1A. (The band, marked as *endo*_{MC}, corresponds to the endogenous *c-myc* gene, which is present in both the normal and the tumour tissue^{3,6}). Each tumour seems to have one TS band co-hybridizing to both probes. (In tumour no. 3, the TS band which joins the REV provirus and the *c-myc* gene co-migrates with the *c-myc* band. The linkage between REV provirus and *c-myc* gene are confirmed by digestion with other restriction enzymes.) Such cross-hybridizations are not due to homology between REV and MC sequences, as the unintegrated CSV DNA (lane R) is not detected by the MC probe. These results suggest that in each tumour, at least one REV provirus is integrated next to the *c-myc* gene (on one of the two homologous chromosomes) and strongly implicate the latter gene in lymphomagenesis.

The REV proviruses in 22 of the 25 tumours characterized are found to be linked to *c-myc* (Fig. 1 and unpublished data). The sizes of the *Eco*RI linkage fragments vary from 11 to 22 kb. The different sizes of TS bands suggest that integrations of the viral DNAs at several sites near the *c-myc* gene are conducive to transformation, a conclusion consistent with previous studies of the LLV-induced tumours^{1-4,18}. A close examination of the TS band patterns, however, reveals that the size variation of some of these bands may also be caused by deletion or structural alteration of the cellular or viral sequences. If the TS bands result from a simple integrative recombination involving the

* To whom correspondence should be addressed.

Fig. 1 The structure of REV proviruses and the *c-myc* gene in lymphoma as analysed using *EcoRI* digestion. Avian bursal lymphomas were induced by injections of end-point purified, helper-independent chicken cyncytial viruses (CSV; a member of the REV family) intra-abdominally into 1-day-old chicks of line 151₁×7₁. After a latent period of ~20 weeks, birds which developed lymphomas were killed and the tumour tissues collected⁵. Both the virus stock and the tumour samples were shown serologically and biochemically to be free of avian leukosis viruses. The DNAs were extracted from the tissues, digested with *EcoRI*, analysed by 0.8% agarose gel electrophoresis and transferred onto nitrocellulose filters, all as previously described¹⁵. Filter hybridizations were carried out in 50% formamide and 3× SSC (1×SSC=0.15 M sodium chloride/0.015 M sodium citrate) at either 35°C (panels A–C, E) or 45°C (panel D). The radiolabelled hybridization probes were prepared by nick-translation¹⁶ of the following three pBR322-based DNA clones: a, pSNV-*Sall* 60B (a gift from H. Temin). This clone, which carries the entire genomic sequence (7.7 kb) of spleen necrosis virus (SNV)²¹, cloned in the *Sall* site of the pBR322 vector, was used to detect the REV sequence in tumour DNAs (the REV probe). SNV and CSV are genetically closely related and their genomes share ≥80% sequence homology (ref. 14 and M.R.N.-D. and H.-J.K., unpublished data).

b, pMC-*Pst* (a gift from D. Sheiness and J. M. Bishop). This clone, which principally carries the oncogene sequence of MC-29 virus²², was used to detect the *c-myc* gene (the MC probe). c, pSRC-*pvuII*E (a gift from W. L. Delorbe and H. E. Varmus). This clone, which carries 0.8 kb of the oncogene sequence of Rous sarcoma virus²³, was used to detect the cellular *src* gene in chicken chromosome (the *src* probe). Panels A, B, D, E, *EcoRI*-digested bursal DNA samples from an uninfected bird (lane U) or from tumorous birds (lanes 1–6). Panel E, the same filter as panel D, after removing the radioactive REV sequences by heating at 90°C in TE buffer (5 mM Tris pH 6.5, 1 mM EDTA), were hybridized with the *src* probe as described above. Panel C, the linear, unintegrated CSV isolated from the cytoplasm of chicken embryo fibroblasts infected with the CSV viruses 48 h earlier by the procedures of Shank *et al.*¹⁵. Lanes a and b represent DNA samples before or after *EcoRI* (Miles) digestion. All except panel F are autoradiograms of the filters after hybridization with the REV, MC or SRC probes (as indicated at the top of the gel). Panel F shows the ethidium bromide (EtBr)-stained DNA pattern of the gel used to prepare the filter shown in panels D and E. The molecular size markers (in kilobases, kb) shown on the side of the gels correspond to the migration pattern of the *EcoRI*-digested phage λ DNA (shown in lane λ of panel F), which was included in each gel run as an internal size standard.

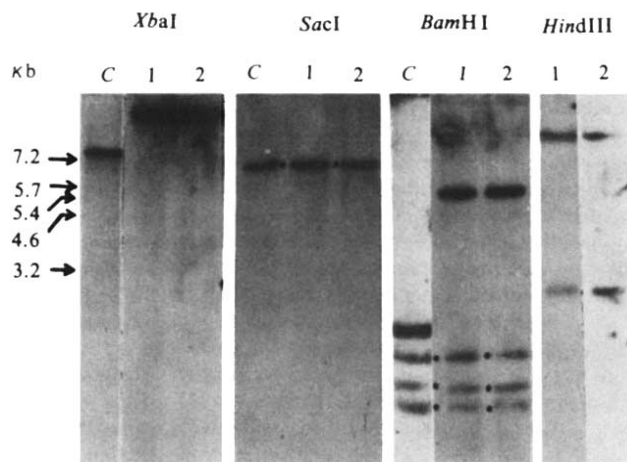
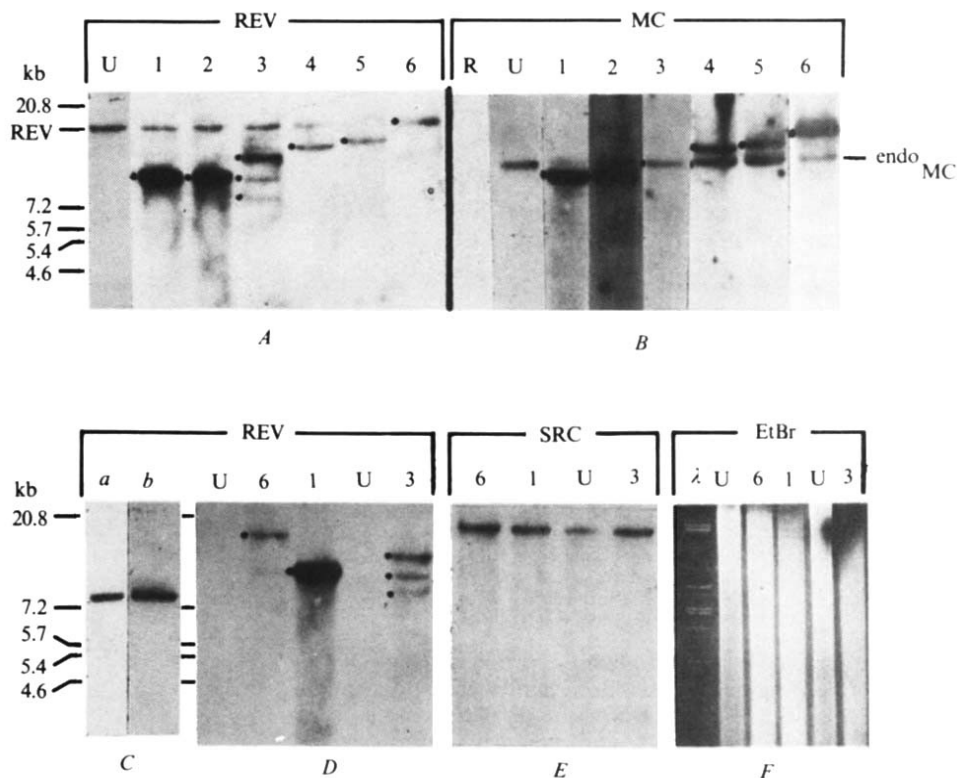


Fig. 2 Comparison of the structure of the REV proviruses in tumours no. 1 and no. 2. Chromosomal DNAs extracted from tumour no. 1 (lane 1), tumour 2 (lane 2) or the cytoplasmic unintegrated CSV DNA (lane C) were digested with restriction endonucleases (Biolabs and Biotech) as indicated at the top of the gels. The samples were analysed in 0.8% agarose gel and the Southern-filter hybridizations with the REV probe were conducted at 45°C as described for Fig. 1D. Among the enzymes used here, *XbaI* has no cleavage site, *HindIII* has one cleavage site and *SacI* and *BamHI* have multiple cleavage sites in the CSV DNA (ref. 8 and R.A.S. and H.-J.K., unpublished data).

bands are actually smaller than the *c-myc* fragment. This result indicates that a deletion or structural alteration (leading to the generation of a new *EcoRI* cleavage site) must have occurred. Further digestion analyses of the proviruses of samples 1 and 2 using other enzymes, including those which have multiple cleavage sites in viral DNA to generate the internal fragments (for example, *SacI* and *BamHI*), are shown in lanes 1 and 2 of Fig. 2. In all cases, identical patterns are displayed by these two proviruses, confirming the identity of the two proviruses. In addition, these proviruses seem to carry the major *BamHI* and *SacI* internal fragments (indicated by dots), comigrating with those generated from the cytoplasmic unintegrated viral DNA (Fig. 2, lane C). This indicates that the majority of viral sequences are present in these proviruses. Thus, the structural alterations which gave rise to the 'shortened' 11-kb TS band probably involve the surrounding cellular sequence.

The other interesting feature of the results shown in Fig. 1D is the unusual intensity of the hybridization signal of some of the TS bands, especially the 11-kb band in sample 1. The total amount of the chromosomal DNA loaded on each lane, as judged by the ethidium bromide stains, is comparable (Fig. 1F). Furthermore, when the same samples were hybridized with a cloned DNA carrying a portion of *v-src* (the transforming gene of Rous sarcoma virus), the intensity of the cellular *src* genes seem to be similar in all lanes. This confirms that the amount of chromosomal DNA of lane 1 is not appreciably different from that of other lanes. The high intensity of the 11-kb TS band in sample 1 is revealed by both REV and MC probes (Fig. 1A, B). This observation, together with the larger size of this band relative to the unintegrated viral DNA, suggests that amplification of the REV provirus and its surrounding cellular

REV DNA and the *c-myc* gene without other structural alterations, the *EcoRI*-TS band should have a size equal to the sum of the viral DNA and endo_{MC} fragment. We found that this is not the case; especially obvious are samples 1 and 2 where the TS

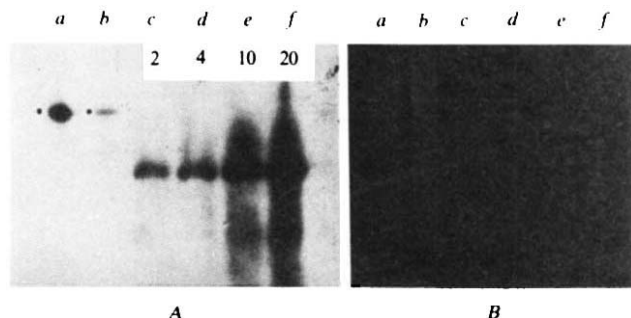


Fig. 3 Quantification of the CSV sequences in tumour no. 1. Panel A, 30 µg each of the *Eco*RI-digested chromosomal DNA from tumours no. 1 (lane a) and no. 8 (lane b) were analysed on a 0.8% agarose gel, then hybridized with the REV probe. The DNA concentration was determined by diphenylamine assay¹⁷. The cloned pSNV-Sal-60B DNA (lanes c-f) was used as a quantitation standard. The DNA insert which carries the 7.7-kb viral sequence was purified from the pBR322 vector by *S*all cleavage and isolated from a preparative agarose gel. Lanes c, d, e and f contain 0.23, 0.46, 1.15 and 2.30 ng each of the purified SNV DNA insert. Based on a diploid genome size of 2×10^9 bases for chicken, these amounts of SNV DNA would give a 2, 4, 10 and 20 copies per cell equivalent of CSV sequences in 30 µg of chicken DNA. The weak, lower band in lanes c-f represents the contaminating pBR322 vector DNA. Panel B, the ethidium bromide-stained gel of panel A.

sequences have occurred. To obtain a further estimate of the amount of CSV sequences present in tumour no. 1, a titration experiment using the pSNV DNA (the plasmid clone used to make REV probe; see Fig. 1 legend) as quantitation standards was carried out (Fig. 3A). Thirty microgrammes each of tumours no. 1 (lane a) and no. 8 (lane b) DNA, digested with *Eco*RI, were loaded on an agarose gel together with varying amounts of the purified 7.7-kb insert of the pSNV DNA (lanes c-f). The quantities of the cloned DNA in lanes c-f were such that they represent, respectively, 2, 4, 10 and 20 copies per cell equivalent of CSV sequence in 30 µg of chicken DNA. We included tumour no. 8 DNA as a control, because it has an *Eco*RI-TS band similar in size to that of tumour no. 1, but has a usual, non-amplified intensity. (The ethidium bromide-stained gel is shown in Fig. 3B.) When the TS bands in lanes a and b were traced densitometrically and their intensity compared with that of the standards in lanes c-f, it was found that while tumour no. 8 acquired only one copy of the CSV provirus, tumour no. 1 appears to carry approximately eight copies of the proviral sequences.

The observation that the REV provirus and its adjacent cellular sequences may have undergone structural alteration in some of the tumours is intriguing in the light of the recent findings that retroviral DNA has a transposable-element-like structure⁸⁻¹² which behaves like prokaryotic transposable-elements in showing, for example, deletion formation and insertional mutagenesis^{13,20}. It is not clear how such structural alterations contribute to the transformation process, if indeed they do. Detailed structural analysis of the molecularly cloned TS bands should provide insight into the molecular mechanism of the recombination and their possible functions in transformation. Finally, the endogenous REV sequences detected in Fig. 1A merits discussion; such an *Eco*RI fragment has not previously been reported, although earlier liquid hybridization studies indicated a low level of homology (10–20%) between REV and the normal chicken chromosome¹⁴. We do not know whether this endogenous REV sequence is specific for the inbred line used here or is ubiquitous in chickens. We are now conducting restriction enzyme digestion analysis of the endo-REV to determine the structural relationship between the endogenous and exogenous viral genomes.

We thank Drs J. M. Bishop, B. Delorbe, D. Sheiness, H. Temin and H. E. Varmus for providing the cloned DNAs for the syntheses of radioactive probes, Drs H. Temin, H. E. Varmus, M. Fluck, F. Rottman, E. Fritsch and E. J. Smith for critical reviews of the manuscript, and Ms S. Uselton for secretarial assistance. This work was supported in part by USPHS grant CA 2479-01, by NSF grant PCM-8021839 and by a grant from Michigan State University Foundation.

Received 6 April; accepted 9 October 1981.

- Hayward, W., Neel, B. G. & Astrin, S. M. *Nature* **290**, 475–480 (1981).
- Neel, B. G. *et al. Cell* **23**, 323–334 (1981).
- Fung, Y. K., Fadley, A. M., Crittenden, L. B. & Kung, H. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3418–3422 (1981).
- Payne, G. S. *et al. Cell* **23**, 311–322 (1981).
- Witter, R. L. & Crittenden, L. B. *Int. J. Cancer* **23**, 673–678 (1979).
- Sheiness, D. K. *et al. Virology* **105**, 415–424 (1980).
- Neiman, P., Payne, L. N. & Weiss, R. A. *J. Virol.* **34**, 178–186 (1981).
- Shimotohno, K., Mizutani, S. & Temin, H. M. *Nature* **285**, 550–554 (1980).
- Temin, H. M. *Cell* **21**, 559–560 (1980).
- Majors, J. M. & Varmus, H. E. *Nature* **289**, 253–258 (1981).
- Dhar, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3937–3941 (1980).
- Ju, G. & Skalka, A. M. *Cell* **22**, 379–386 (1980).
- Majors, J. E. *et al. Cold Spring Harb. symb. quant. Biol.* **45** (in the press).
- Kang, C. & Temin, H. M. *J. Virol.* **35**, 888–894 (1973).
- Shank, P. R. *et al. Cell* **15**, 1383–1395 (1978).
- Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184–1188 (1975).
- Putman, E. W. *Meth. Enzym.* **3**, 99 (1957).
- Payne, G. S., Bishop, J. M. & Varmus, H. E. *Nature* (in the press).
- Cooper, G. & Neiman, P. *Nature* **292**, 857–858 (1981).
- Varmus, H. E., Quintrell, N. & Ortiz, S. *Cell* **25**, 23–36 (1981).
- O'Rear, J. J. *et al. Cell* **20**, 423–430 (1981).
- Vennstrom, B. *et al. J. Virol.* **39**, 625–631 (1981).
- Delorbe, W. J. *et al. J. Virol.* **36**, 50–61 (1981).

An isolated pseudogene related to the 5S RNA genes in *Neurospora crassa*

Eric U. Selker[‡], Stephen J. Free[†],
Robert L. Metzenberg[†] & Charles Yanofsky^{*}

^{*} Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

[†] Department of Physiological Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA

DNA segments which do not code for functional products but are similar in sequence to DNA of known function are called 'pseudogenes'. Most pseudogenes discovered so far are related to members of gene families: 5S RNA (5S) genes in *Xenopus*¹⁻³, globin genes in various mammals⁴⁻⁶, actin genes in *Dictyostelium*⁷ and small nuclear RNA genes in man⁸. Presumably, in the absence of strong selective forces, redundant genes drift apart and occasionally become nonfunctional. Of course, mechanisms which maintain homogeneity in multigene families should tend to prevent the occurrence of pseudogenes. There is ample evidence for 'horizontal' or 'concerted' evolution in tandemly repeated gene families⁹⁻¹¹. The 5S pseudogenes of *Xenopus laevis*, of which there is a copy in every tandemly repeated unit, may be a rare example of a defective gene not eliminated from the family by unequal crossing-over or gene conversion events. We have recently found evidence for limited concerted evolution in a dispersed gene family, the 5S genes of *Neurospora crassa*¹²: both identical and highly divergent 5S coding regions were discovered. We report here that we have found a pseudogene related to the *Neurospora* 5S genes. This pseudogene, N5SP1, contains a segment of DNA almost identical to the first 50 nucleotides of the major species of *Neurospora* 5S RNA. The sequences flanking this region are totally different from sequences found within or adjacent to full-length 5S genes. Hybridization experiments suggest that N5SP1 is not transcribed. The existence of this pseudogene is consistent with our transposition model for horizontal evolution in dispersed gene families¹².

N5SP1 is contained on a plasmid designated pMF4. It was identified in a collection of *Neurospora* DNA fragments cloned into pBR322 (ref. 13) by colony hybridization¹⁴ using labelled 5S RNA as probe. The first indication that pMF4 does not contain a complete 5S gene came from Southern¹⁵ hybridization

[‡] Present addresses: Max-Planck-Institut für Züchtungsforschung, D5000, Köln 30, FRG (E.U.S.); Department of Biological Sciences, State University of New York, Buffalo, New York 14260, USA (S.J.F.).

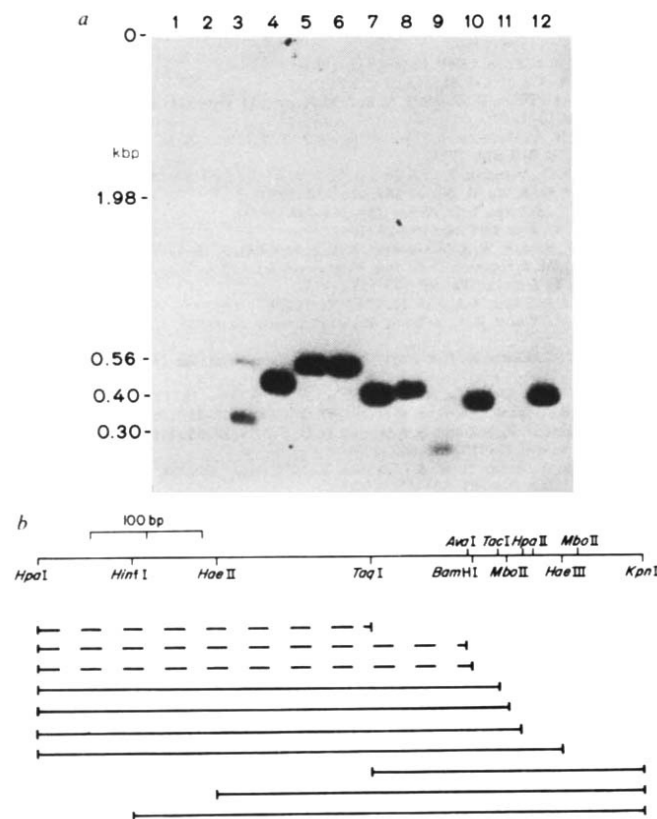


Fig. 1 Hybridization of 5S RNA to restriction endonuclease digests of pMF4 DNA. *a*, Plasmid DNA (2 µg) was digested with *KpnI* and *HpaI*, together with *AvaI* (lane 1), *BamHI* (lane 2), *HaeII* (lane 3), *HaeIII* (lane 4), *HhaI* (lane 5), *HincII* (lane 6), *HpaII* (lane 7), *HinfI* (lane 8), *TaqI* (lane 9), *TacI* (lane 10), *Sau3A* (lane 11), or *MboII* (lane 12). The resulting DNA fragments were resolved on a 1.5% agarose gel in Tris-EDTA-acetate-NaCl buffer²³ and transferred to a nitrocellulose filter according to the method of Southern¹⁵. The filter was then hybridized with ³²P-labelled 5S RNA¹² at 65 °C overnight in 5×SSC, 0.1% SDS. It was then washed extensively with 5×SSC, blot-dried and autoradiographed. The positions of size standards (in kilobase pairs, kbp) and of the origin (O) are indicated. *HhaI* (lane 5) and *HincII* (lane 6) did not cleave the *KpnI*-*HpaI* DNA fragment and the *HaeII* digestion (lane 3) did not go to completion. The DNA fragments which hybridized to 5S RNA in lanes 3 and 9 were too small to transfer effectively to nitrocellulose and therefore produced weak bands. *b*, Restriction map of *KpnI*-*HpaI* segment of pMF4 (data not shown) and interpretation of hybridization results. Solid horizontal lines indicate DNA fragments which hybridized to 5S RNA and broken lines indicate fragments which failed to hybridize. Fragments smaller than the *TaqI*-*KpnI* fragment did not transfer efficiently, and thus cannot be scored. Note that the *HpaI*-*TacI* fragment hybridized well with 5S RNA, whereas the 19-nucleotide shorter *HpaI*-*BamHI* (or *Sau3A*) fragments did not hybridize at all. This is presumably because the *BamHI* cut is in the middle of the largest uninterrupted stretch of homology to 5S RNA.

experiments. Restriction digests of pMF4 DNA were fractionated by agarose gel electrophoresis, transferred to a nitrocellulose filter and hybridized with ³²P-labelled 5S RNA. The results of this experiment showed that 5S RNA hybridizes to pMF4 DNA within a 550-base pair (bp) *KpnI*-*HpaI* fragment located near the middle of the 5,400-bp *Neurospora* DNA insert (data not shown). Surprisingly, digestion of pMF4 DNA with *BamHI*, which cleaves the *KpnI*-*HpaI* fragment, eliminates all subsequent hybridization with 5S RNA. Functional *Neurospora* 5S genes contain a *BamHI* site at nucleotide 32 in the coding region, but DNA cleaved at this site still hybridizes to 5S RNA¹². These findings suggest that a *BamHI* site in pMF4 is in a region of incomplete 5S homology. The DNA segment homologous to 5S RNA was mapped by Southern hybridizations of pMF4 DNA digested with *KpnI* and *HpaI* together with other restriction endonucleases. The results (Fig. 1a) are consistent with the idea that the region of 5S homology is near the *BamHI* site (Fig. 1b). This region was sequenced as indicated in Fig. 2a and compared with sequences of two *Neurospora* 5S genes (Fig. 2b). As expected, the *BamHI* site is in a region homologous to 5S RNA. A ~50-nucleotide segment

corresponds to the first ~40% of 5S RNA. With the exception of a three-nucleotide insertion, a single nucleotide insertion and two single nucleotide substitutions, this region matches the major 5S RNA sequence found in *Neurospora*. Outside this segment, there is no significant homology with DNA in or around the *Neurospora* 5S genes. *Neurospora* mitochondrial DNA does not hybridize to pMF4 (data not shown), ruling out the possibility that the cloned DNA encodes a mitochondrial RNA similar to cytoplasmic 5S RNA.

The Southern hybridization experiments and DNA sequence analysis suggests that pMF4 does not encode full-length 5S RNA. S₁ nuclease mapping¹⁶ was performed to verify this (Fig. 3). pMF4 DNA 3'- or 5'-labelled at the *BamHI* site in the region of 5S homology was hybridized with excess 5S RNA in R-looping¹⁷ conditions. The single-stranded regions were digested with S₁ nuclease and the resulting hybrids denatured, fractionated on a denaturing polyacrylamide gel and autoradiographed. DNA fragments protected by RNA were detected on one side of the *BamHI* site (corresponding to the 5' portion of 5S RNA) but not on the other, confirming that pMF4 does not encode a 5S RNA. For the authentic 5S clones pKD51 and pKD52, fragments corresponding to both the 5' and 3' portions were evident (Fig. 3).

To determine whether the presumptive pseudogene encodes RNA of a different size from 5S RNA, total *Neurospora* RNA fractionated on native (2.1% polyacrylamide, 0.5% agarose) or denaturing (7 M urea, 10% polyacrylamide) gels was transferred to diazobenzylxymethyl paper^{18,19}, and probed with labelled pMF4 DNA. In each case, the only RNA species detected corresponded to 5S RNA, so it seems likely that the 5S-like DNA segment in pMF4 is a 5S pseudogene, which we designate N5SP1.

The sharp break in the sequence homology between N5SP1 and 5S RNA presumably reflects a DNA rearrangement which replaced the 3' portion of the ancestral 5S gene with foreign DNA. This substitution separated the 5' portion of the 5S gene from the central region which, at least in *X. laevis* 5S RNA genes²⁰⁻²², is necessary for transcription initiation. A gene inactivated by such a rearrangement might be expected to accumulate subsequent mutations. The small insertions and point changes in N5SP1, relative to active 5S genes, presumably are examples of this. Other pseudogenes which have been discovered¹⁻⁸ also contain similar mutational changes.

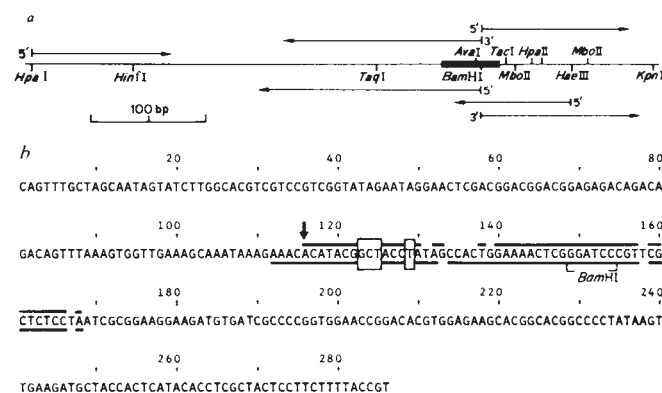


Fig. 2 DNA sequence of *Neurospora* 5S pseudogene, N5SP1. *a*, Restriction map with horizontal arrows illustrating sequencing runs on the two strands of DNA (above and below map). The lengths of the arrows indicate the extent to which the sequence could be read with confidence. DNA fragments were labelled at 5' termini by phosphorylation using [γ -³²P]ATP and T4 polynucleotide kinase²¹ or at 3' termini by 'filling in' using [α -³²P]dGTP and *Escherichia coli* DNA polymerase I^{24,25}. They were then sequenced according to the methods of Maxam and Gilbert²⁶. Reaction products were resolved on 0.35-mm thick 20% or 8% polyacrylamide gels containing 7 M urea²⁷. The heavy line indicates the region of homology to 5S RNA. *b*, DNA sequence of one strand in the vicinity of the *BamHI* site in N5SP1. The horizontal lines above and below the sequence show the extent of homology with *Neurospora* β and α 5S genes on pKD51 and pKD52, respectively¹². Boxes enclose two small insertions relative to the 5S genes. The arrow shows the presumed position of transcription initiation in the 5S genes.

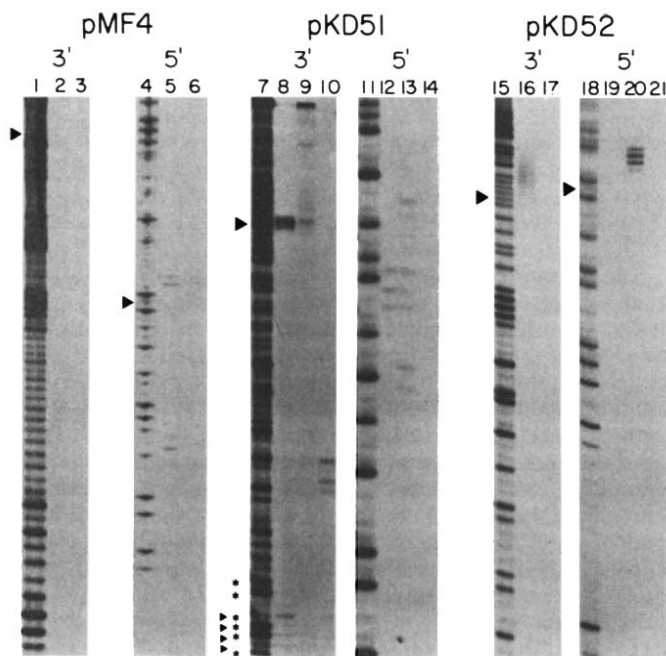


Fig. 3 S_1 nuclease mapping of 5S RNA homology to pseudogene clone pMF4 and authentic 5S clones pKD51 and pKD52. The *Bam*HI ends of DNA fragments *Kpn*I-*Bam*HI (lanes 1-3) and *Pst*I-*Bam*HI (lanes 7-10 and 15-17) were labelled at their 3' ends (see Fig. 2). The *Bam*HI ends of the fragments *Taq*I-*Bam*HI were labelled at their 5' ends (lanes 4-6, 11-14 and 18-21). The labelled DNA was then used for the G > A DNA sequencing reaction²⁶ (lanes 1, 4, 11, 15, 18) or the A > C reaction (lane 7), or hybridized in R-looping conditions (10 μ l containing 0.4 M NaCl, 0.04 M PIPES pH 6.4, 1 mM EDTA and deionized 80% formamide; 55 °C for the 5' fragments and 65 °C for the 3' fragments), with (lanes 2, 5, 8, 9, 12, 13, 16, 19, 20) or without (lanes 3, 6, 10, 14, 17, 21) a ~100-fold excess (0.1 μ g) of 5S RNA. After 16 h, the samples were chilled, diluted 10-fold in 0.25 M NaCl, 0.03 M sodium acetate pH 4.6 and 100 μ g ml⁻¹ denatured salmon sperm DNA, and treated with S_1 nuclease (Boehringer) at 1 U μ l⁻¹ (lanes 2, 3, 5, 6, 8, 10, 13, 14, 17, 20, 21) or 10 U μ l⁻¹ (lanes 9, 12, 16, 19) at 22 °C for 20 min. Finally, the nucleic acids were ethanol-precipitated, rinsed with 70% ethanol, dried and resuspended in 5-10 μ l deionized 99% formamide containing xylene cyanol and bromophenol blue and fractionated on 8% or 20% polyacrylamide sequencing gels²⁷ next to the same DNA which had undergone partial cleavage in a DNA sequencing reaction. Large arrowheads indicate the expected migration of the fragments from the *Bam*HI site to 3' or 5' ends of the 5S coding regions of pKD51 and pKD52. Small arrowheads mark the positions of faint bands in lane 8 where bulk 5S RNA fails to protect completely pKD51. This is near a region (*) of recognized¹² non-identity between pKD51 and pKD52.

If a defective 5S gene such as N5SP1 were part of a tandemly repeated multigene family, it could be eliminated by unequal crossing-over. Alternatively, it could be passively amplified as apparently occurred in the case of the *Xenopus laevis* 5S pseudogene^{1,2}. In dispersed gene families, defective members cannot be lost by unequal crossing-over. However, there is evidence for limited concerted evolution among the dispersed *Neurospora* 5S genes¹². We pointed out that gene conversion or gene transposition, together with normal recombination processes, should have a homogenizing effect in a family of dispersed, repeated genes¹². A feature of our model for concerted evolution by conservative transposition is that genes which fail to jump are immune to the homogenizing mechanism. The *Neurospora* 5S pseudogene may be an example of a defective gene which no longer participates in concerted evolution. Presumably it will continue to drift until it either becomes functional in some new way, or is deleted.

This work was supported by NIH grant GM 09738 and NSF grant PCM 77-24333 to C.Y. and NIH grant GM 08995 to R.L.M. C.Y. is a Career Investigator of the American Heart Association.

Received 5 July; accepted 7 October 1981.

1. Jacq, C., Miller, J. R. & Brownlee, G. G. *Cell* **12**, 109-120, (1977).
2. Miller, J. R., Cartwright, E. M., Brownlee, G. G., Federoff, N. V. & Brown, D. D. *Cell* **13**, 717-725 (1978).

3. Korn, L. J. & Brown, D. D. *Cell* **15**, 1145-1156 (1978).
4. Proudfoot, N. J., Shander, M. H. M., Manley, J. L., Gefter, M. L. & Maniatis, T. *Science* **209**, 1329-1336 (1980).
5. Leder, P. *et al. Science* **209**, 1336-1342 (1980).
6. Jeffreys, A. J. *et al. Cell* **21**, 555-564 (1980).
7. Firtel, R. A., Timm, R., Kimmel, A. R. & McKeown, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6206-6210 (1979).
8. Denison, R. A., Van Arsdell, S. W., Bernstein, L. B. & Weiner, A. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 810-814 (1981).
9. Brown, D. D., Wensink, P. C. & Jordan, E. J. *molec. Biol.* **63**, 57-73 (1972).
10. Szostak, J. W. & Wu, R. *Nature* **284**, 426-430 (1980).
11. Klein, H. L. & Petes, T. D. *Nature* **289**, 144-148 (1981).
12. Selker, E. U. *et al. Cell* **24**, 819-828 (1981).
13. Free, S. J., Rice, P. W. & Metzberg, R. L. *J. bact.* **137**, 1219-1226 (1979).
14. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3965 (1975).
15. Southern, E. J. *molec. Biol.* **98**, 503-518 (1975).
16. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
17. Thomas, M., White, R. L. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2294-2298 (1976).
18. Alwine, J. C., Kemp, D. J. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5350-5354 (1977).
19. Stellwag, E. J. & Dahlberg, A. E. *Nucleic Acids Res.* **8**, 299-318 (1980).
20. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. *Cell* **19**, 13-25 (1980).
21. Bogenhagen, D. F., Sakonju, S. & Brown, D. D. *Cell* **19**, 27-35 (1980).
22. Sakonju, S. *et al. Cell* **23**, 665-669 (1981).
23. Dugaiczky, A., Boyer, H. W. & Goodman, H. M. *J. molec. Biol.* **96**, 171-184 (1975).
24. Wu, R. J. *molec. Biol.* **51**, 501-521 (1970).
25. Selker, E. & Yanofsky, C. *J. molec. Biol.* **130**, 135-143 (1979).
26. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
27. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107-110 (1978).

Identification of the *Escherichia coli* *recB* and *recC* gene products

Ian D. Hickson & Peter T. Emmerson

Department of Biochemistry, The University, Newcastle upon Tyne NE1 7RU, UK

Functional *recB* and *recC* genes are required for normal conjugative genetic recombination in *Escherichia coli*¹⁻³ and mutations in either of these genes lead to deficiency in an ATP-dependent nuclease known as the RecBC DNase or exonuclease V (refs 4-6). Widely different molecular weights have been reported for the putative subunits of this enzyme^{7,8} and genetic and biochemical experiments have left some doubt about whether or not *recB* and *recC* are different genes. We have now cloned the *recB* and *recC* genes separately into the multiple copy number plasmid pAT153 and identified the gene products by transposon inactivation and gel electrophoresis. The products of the *recB* and *recC* genes are proteins of molecular weights (MW) 135,000 and 125,000, respectively.

Mutations in either the *recB* or *recC* genes lead to apparently identical deficiency in genetic recombination, increased radiation sensitivity and decreased post-irradiation DNA degradation^{1,2}. The *recB* and *recC* mutations cannot be distinguished phenotypically but have been assigned to different genes on the basis of genetic complementation tests and co-transduction frequencies with the neighbouring *argA* and *thyA* genes³. However, complex complementation behaviour has been observed among several recombination deficient mutations in this region (Hoekstra *et al.* cited in ref. 9) and the possibility that the observed complementation between *recB* and *recC* mutations may be intragenic cannot be discounted. *In vitro* complementation studies have also been difficult to interpret. The RecBC enzyme has been partially purified and dissociated by high salt concentrations into two subunits of MW 60,000 and 170,000, as determined by sedimentation in glycerol gradients⁷. It was found that the 170,000-MW protein would complement extracts from both *recB* and *recC* mutant strains whereas the 60,000-MW protein complemented neither⁷. These results suggest that the 170,000-MW protein is either a single polypeptide encoded by one gene of which *recB* and *recC* are mutations, or that it is made up of two polypeptides encoded by *recB* and *recC* which are not dissociated by high salt concentrations. However, in other work⁸, the enzyme has been purified to about 90% homogeneity and preparations were found, by SDS-polyacrylamide gel electrophoresis, to consist mainly of two protein subunits of 128,000 and 140,000 MW.

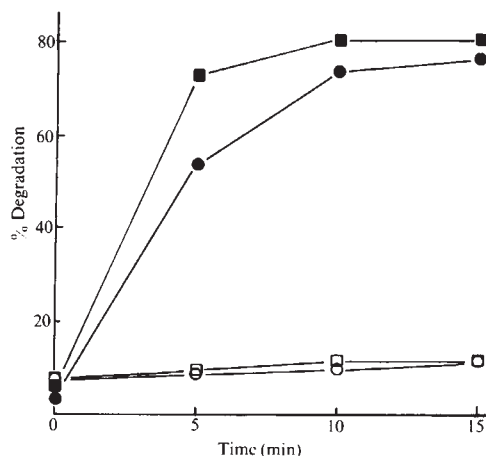


Fig. 1 ATP-dependent DNase activity in cell extracts. ○, AB2470 (*recB*); ●, AB2470 (*recB*) harbouring pPE230 (*recB*⁺); □, AB3058 (*recC*); ■, AB3058 (*recC*) harbouring pPE220 (*recC*⁺). Cell lysates were assayed²⁴ for ability to degrade ³H-labelled phage λ DNA to trichloroacetic acid-soluble material in the presence of ATP. Less than 10% degradation occurred in extracts of all four strains in the absence of ATP. Results with the strain AB1157 (*rec*⁺) were similar to those with AB2470 (pPE230) (data not shown).

In work to be published elsewhere, we first cloned a *Bam*HI fragment of chromosomal DNA which included *recC*⁺ into a phage λ vector to produce a recombinant phage λ*dreC* which was then used to derive the phage λ*dreBC* by secondary site integration at *recC* followed by aberrant excision (I.D.H., K. E. Atkinson and P.T.E., in preparation). We then subcloned the *recB* and *recC* genes from this phage DNA separately into the *Bam*HI site of the plasmid vector pAT153 (Ap^r, Tc^r) and transformed AB2470 (*recB*21) and AB3058 (*recC*22). Colonies resistant to both ampicillin and mitomycin C were selected and screened for those that were sensitive to tetracycline and resistant to UV light. The plasmids pPE230 (*recB*) and pPE220 (*recC*), obtained in this way, when transformed into AB2470 and AB3058, restored wild-type levels of RecBC

enzyme activity (Fig. 1) and UV resistance (Fig. 2). Restriction maps of these plasmids are shown in Fig. 3.

To identify the products of the *recB* and *recC* genes, we inserted the γδ (ref. 10) sequence of the sex factor F from F'*lac* into the genes and examined extracts of 'maxicells' for altered proteins^{11,12}. The strain NH4104 (F'*lac*, Str^r) harbouring either pPE220 or pPE230 was mated with AB3058 (*recC*22) or AB2470 (*recB*21), respectively, and ampicillin-resistant, streptomycin-resistant exconjugants were selected. This selection is for plasmids (which carry Ap^r) transferred on conjugation because they have acquired γδ insertions from the F DNA. Colonies were then picked and screened for those which were UV-sensitive because the inserted transposable element had inactivated the *recB* or *recC* gene on the plasmid, which was therefore no longer able to complement the chromosomal mutation. Plasmid DNA was isolated¹³ to confirm that the γδ sequence was inserted into the original plasmid, as manifested by an increase in size.

The proteins encoded by pPE220 and pPE230 were examined by subjecting labelled 'maxicell' extracts to SDS-polyacrylamide gel electrophoresis and fluorography. The results (Fig. 4) show that the plasmid pPE220 (*recC*) encodes a polypeptide of MW 125,000 (track 6) that is missing in plasmids in which the *recC* gene is inactivated by γδ (tracks 8–10). This polypeptide is present in a plasmid which carries a γδ insert in a gene other than *recC* (track 7). Similarly, the plasmid pPE230

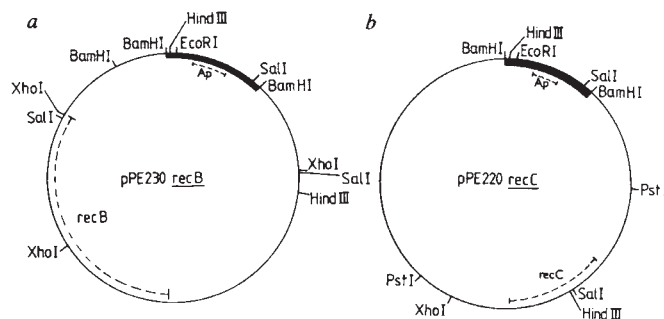


Fig. 3 Restriction maps of the recombinant plasmids. *a*, pPE230 (*recB*); *b*, pPE220 (*recC*). Vector DNA is drawn with a bold line. The inserts are ~17.5 and 21.5 kilobases in pPE230 and pPE220, respectively. Restriction analysis of the plasmids and of their derivatives in which γδ is inserted into the *recB* and *recC* genes shows that these genes are located within the regions indicated.

(*recB*) encodes a polypeptide of MW 135,000 (track 5) that is absent in plasmids in which γδ is inserted into the *recB* gene (tracks 1–3), but present in a plasmid in which it is inserted at a site other than the *recB* gene (track 4). Insertion of γδ into a gene results in termination of transcription in five out of the six possible reading frames^{12,14}, giving rise to truncated peptides. Truncated peptides can be seen in tracks 9 (115,000 MW) and 10 (108,000 MW). Proteins of MWs 125,000 and 135,000 are also synthesized in heavily UV-irradiated host cells infected with λ*dreBC* while a protein of MW 125,000 but not one of 135,000 is synthesized in similar experiments with the phage λ*dreC* which lacks *recB* (I.D.H., K. E. Atkinson and P.T.E., in preparation).

We conclude that the *recB*⁺ gene product is a protein of MW 135,000, the *recC*⁺ gene product is a protein of 125,000, and that *recB* and *recC* are different genes. Further evidence that these are different genes is provided by restriction analysis (Fig. 3), which fails to reveal any similarities between the restriction patterns of the inserts, and the complementation analysis (Fig. 2), which shows that there is a very sharp distinction between the abilities of pPE230 (*recB*) and pPE220 (*recC*) to complement *recB* and *recC* mutations. Our results do not support the conclusion⁷ that the *recB* and *recC* genes are both required to produce one of the subunits of the RecBC enzyme, but suggest that the two peptides of MWs 140,000 and 128,000 found⁸ in

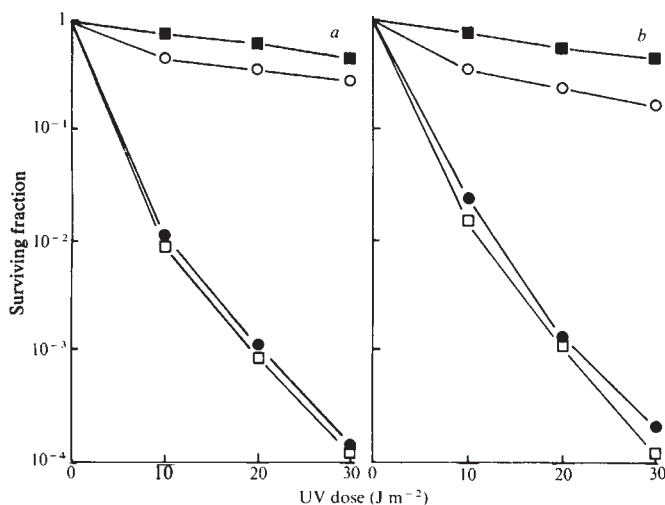


Fig. 2 UV-survival curves showing complementation of the *recB* mutation by pPE230 (*recB*⁺) and the *recC* mutation by pPE220 (*recC*⁺). *a*, ○, AB3058 (*recC*22) harbouring pPE220 (*recC*⁺); ●, AB3058 (*recC*22) harbouring pAT153; ■, AB1157 (*rec*⁺) harbouring pAT153. *b*, ○, AB2470 (*recB*21) harbouring pPE230 (*recB*⁺); ●, AB2470 (*recB*21) harbouring pAT153; ■, AB1157 (*rec*⁺) harbouring pAT153. The bacteria were grown to mid-exponential phase, collected, serially diluted in buffer, spread on nutrient agar plates, irradiated with a low-pressure mercury germicidal lamp and incubated overnight at 30 °C.



Fig. 4 35 S-methionine labelled proteins produced in maxicells by the plasmids pPE230 (*recB*⁺) and pPE220 (*recC*⁺) and their derivatives in which $\gamma\delta$ is inserted in the *recB* or *recC* gene ($\gamma\delta$ 1–3) or in a gene other than *recB* or *recC* ($\gamma\delta$ 0), Tracks: 1, pPE230- $\gamma\delta$ 3; 2, pPE230- $\gamma\delta$ 2; 3, pPE230- $\gamma\delta$ 1; 4, pPE230- $\gamma\delta$ 0 (*recB*⁺); 5, pPE230 (*recB*⁺); 6, pPE220 (*recC*⁺); 7, pPE220- $\gamma\delta$ 0 (*recC*⁺); 8, pPE220- $\gamma\delta$ 1; 9, pPE220- $\gamma\delta$ 2; 10, pPE220- $\gamma\delta$ 3. Proteins were separated by electrophoresis on an 8% polyacrylamide gel and visualized by fluorography. Molecular weights, determined by reference to unlabelled protein standards followed by staining, are indicated. The standards were RNA polymerase β' (165,000), RNA polymerase β (155,000), phosphorylase *b* (96,000), bovine serum albumin (68,000) and ovalbumin (45,000).

preparations of the RecBC enzyme that were 90% pure were the *recB* and *recC* products, respectively.

Our cloning of the *recB* and *recC* genes individually into a multicopy plasmid may lead to an easier method of isolating the individual subunits of the RecBC enzyme and should permit investigation of their partial properties. Our results do not rule out the possibility that active RecBC DNase includes one or more subunit(s) encoded by a gene or genes other than *recB* and *recC*. However, purification of the individual RecB and RecC subunits will permit studies of the reconstituted enzyme to help answer this question. Recent results show that at physiological ATP concentrations and in the presence of Ca^{2+} the RecBC DNase uses energy derived from ATP hydrolysis to travel through duplex DNA, unwinding the DNA ahead of itself and rewinding it behind^{15–17}. In this regard, the RecBC DNase resembles other enzymes such as the helicases, gyrase and RecA protein which use the energy derived from ATP hydrolysis to modify the physical structure of DNA. Cloning of the *recA* gene^{18–20} and the resultant ease with which the RecA protein could be isolated paved the way for the biochemical experiments which have greatly helped to explain the role of the RecA protein in genetic recombination and DNA repair^{21–23}. The cloned *recB* and *recC* genes may also facilitate the isolation of the RecBC DNase and stimulate further experiments to determine its role in the cell.

We thank Pauline Martin and Linda Hutton for technical assistance. This work was supported by the MRC.

Received 28 July; accepted 9 October 1981.

- Emmerson, P. T. & Howard-Flanders, P. *J. Bact.* **93**, 1729–1731 (1967).
- Emmerson, P. T. *Genetics* **60**, 19–30 (1968).
- Willets, N. S. & Mount, D. W. *J. Bact.* **100**, 923–934 (1969).
- Buttin, G. & Wright, M. *Cold Spring Harb. Symp. quant. Biol.* **33**, 259–269 (1968).
- Oishi, M. *Proc. natn. Acad. Sci. U.S.A.* **64**, 1292–1299 (1969).
- Goldmark, P. J. & Linn, S. *Proc. natn. Acad. Sci. U.S.A.* **67**, 434–441 (1970).
- Lieberman, R. P. & Oishi, M. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4816–4820 (1974).
- Goldmark, P. J. & Linn, S. *J. biol. Chem.* **247**, 1849–1860 (1972).
- Clark, A. J. *Ann. Rev. Microbiol.* **25**, 438–464 (1971).
- Guyer, M. S. *J. molec. Biol.* **126**, 347–365 (1978).
- Sancar, A., Hack, A. M. & Rupp, W. D. *J. Bact.* **137**, 692–693 (1979).
- Sancar, A. *et al. J. molec. Biol.* **148**, 45–62 (1981).
- Birnboim, H. C. & Doly, J. *Nucleic Acids Res.* **7**, 1513–1523 (1979).
- Reed, R. R., Young, R. A., Steitz, J. A., Grindley, N. D. F. & Guyer, M. S. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4882–4886 (1979).
- Rosamund, J., Telander, K. M. & Linn, S. *J. biol. Chem.* **254**, 8646–8652 (1979).
- Telander Muskavitch, K. M. & Linn, S. in *Mechanistic Studies of DNA Replication and Genetic Recombination* (ed. Alberts, B.) 909–918 (Academic, London, 1980).
- Taylor, A. & Smith, G. R. in *Mechanistic Studies of DNA Replication and Genetic Recombination* (ed. Alberts, B.) 901–908 (Academic, London, 1980).

- McEntee, K. & Epstein, W. *Virology* **77**, 306–318 (1977).
- Ogawa, T. *et al. Cold Spring Harb. Symp. quant. Biol.* **43**, 909–915 (1978).
- Sancar, A. & Rupp, W. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3144–3148 (1979).
- Weinstock, G. M., McEntee, K. & Lehman, I. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 126–130 (1979).
- Shibata, T., Cunningham, R. P., DasGupta, C. & Radding, C. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5100–5104 (1979).
- West, S. C., Cassuto, E., Mursallim, J. & Howard-Flanders, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2569–2573 (1980).
- Barbour, S. D. & Clark, A. J. *Proc. natn. Acad. Sci. U.S.A.* **65**, 955–961 (1970).

Sexual activity reduces lifespan of male fruitflies

Linda Partridge & Marion Farquhar

Department of Zoology, University of Edinburgh,
West Mains Road, Edinburgh EH9 3JT, UK

Many theories on the evolution of life histories have assumed a physiological cost of reproduction in terms of reduced lifespan^{1–3}. A cost of increased reproduction in terms of reduced longevity has been established experimentally for females, both as an additive genetic^{4,5} and as a purely phenotypic^{6,7} effect. Such a physiological cost of reproduction has not been demonstrated for males. The cost of sexual activity has been assumed to be relatively small in those species where the only paternal contribution to an offspring is the gamete^{8,9}. Here we show that increasing sexual activity reduces longevity in the male fruitfly (*Drosophila melanogaster*) and hence that there is a significant physiological cost of male sexual activity in a species where the father contributes only gametes to his progeny.

The flies used were an outbred stock collected in Dahomey in 1970. Sexual activity was manipulated by supplying individual males with receptive virgin females at a rate of one or eight virgins per day. The longevity of these males was recorded and compared with that of two control types. The first control consisted of two sets of individual males kept with newly inseminated females equal in number to the virgin females supplied to the experimental males. Newly inseminated females will not usually re-mate for at least 2 days^{10,11} thus they served as a control for any effect of competition with the male for food or space. The second control was a set of individual males kept with no females. There were 25 males in each of the experimental and control groups, and the groups were treated identically in respect of number of anaesthetizations (using CO₂) and provision of fresh food medium.

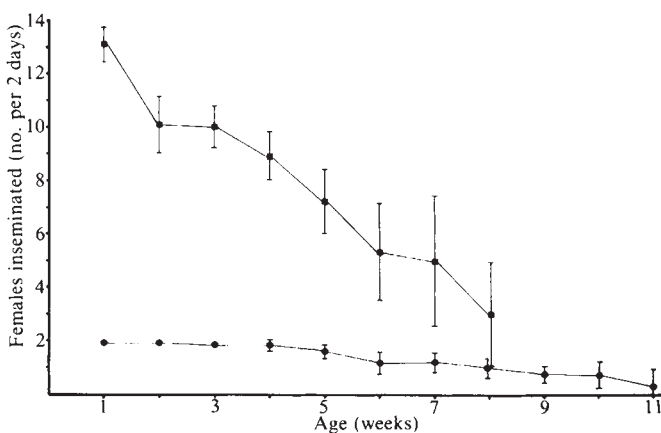


Fig. 1 The relationship between insemination rate (number of females inseminated per 2 days) and age (weeks) for males kept with one virgin female (●) or eight virgin females (■) per day. Error bars are 95% confidence limits.

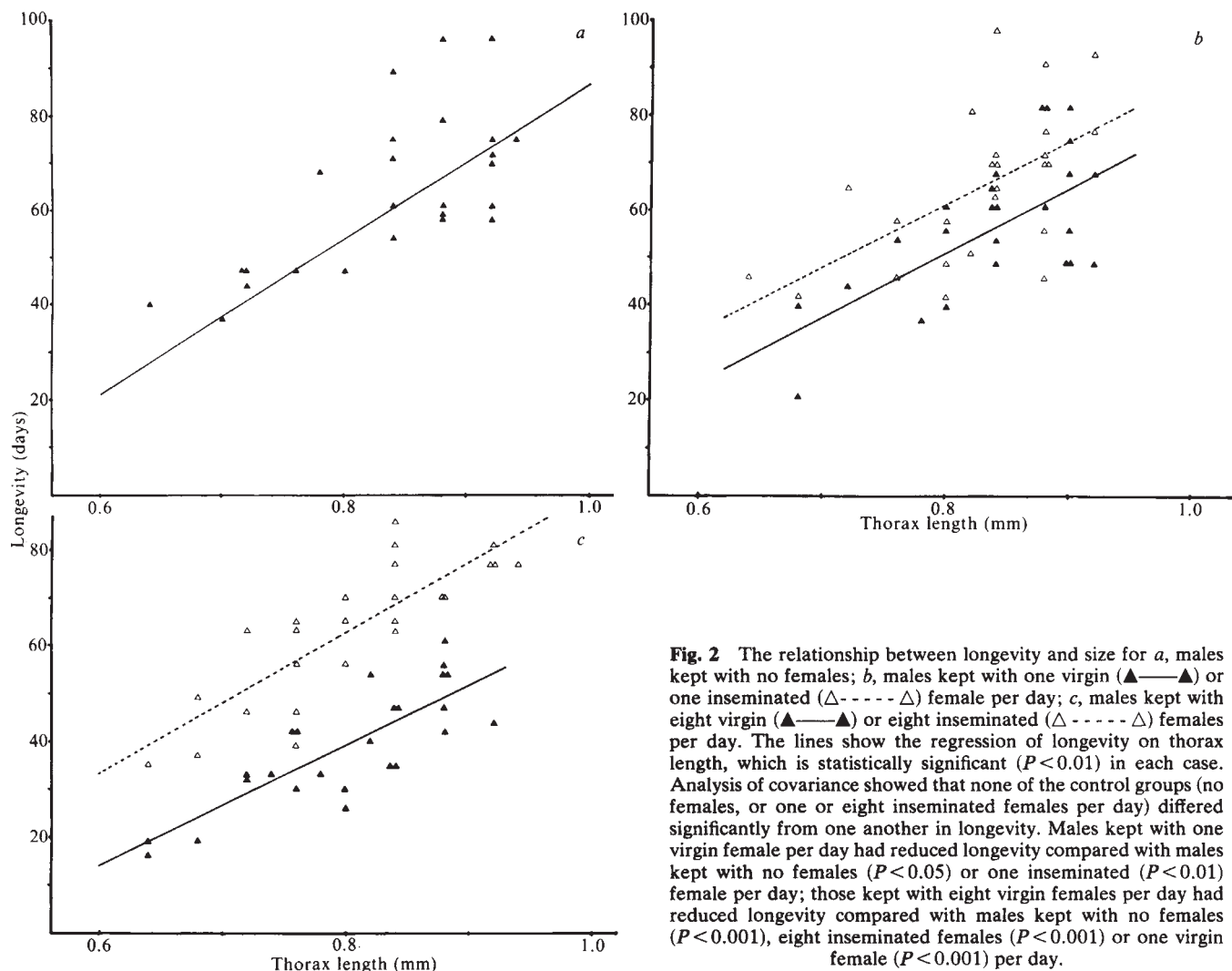


Fig. 2 The relationship between longevity and size for *a*, males kept with no females; *b*, males kept with one virgin ($\blacktriangle$ — $\blacktriangle$) or one inseminated ($\triangle$ — $\triangle$) female per day; *c*, males kept with eight virgin ($\blacktriangle$ — $\blacktriangle$) or eight inseminated ($\triangle$ — $\triangle$) females per day. The lines show the regression of longevity on thorax length, which is statistically significant ($P < 0.01$) in each case. Analysis of covariance showed that none of the control groups (no females, or one or eight inseminated females per day) differed significantly from one another in longevity. Males kept with one virgin female per day had reduced longevity compared with males kept with no females ($P < 0.05$) or one inseminated ($P < 0.01$) female per day; those kept with eight virgin females per day had reduced longevity compared with males kept with no females ($P < 0.001$), eight inseminated females ($P < 0.001$) or one virgin female ($P < 0.001$) per day.

On two days per week throughout the life of each experimental male, the females that had been supplied as virgins to that male were kept and examined for fertile eggs. This gave an estimate of the insemination rate of the two groups of males (Fig. 1). In both groups the insemination rate declined with the age of the male, and the rate was higher for males supplied with eight virgins per day than for those supplied with only one virgin per day. There were no significant differences in insemination rate between the individual males within each experimental group.

In the absence of any sexual activity, the longevity of male fruitflies is associated with their size (Fig. 2*a*). Size was therefore taken into account when examining the effect of sexual activity on longevity (Fig. 2*b, c*). Analysis of covariance showed that: (1) there were no significant differences in longevity between the control groups (median longevity 65 days); (2) the males supplied with one virgin female per day had significantly reduced longevity (median 56 days) compared with males in any control group; (3) males kept with eight virgin females per day had significantly reduced longevity (median 40 days) compared with males kept with one virgin female per day, and control males. These results show that male sexual activity reduces longevity, and that this effect is more marked for a higher level of sexual activity.

Physiological costs of particular activities have generally been discussed in terms of the diversion of nutrients into these activities at the expense of others¹². In our experiment, energetic costs of sexual activity would have included the production of sperm and seminal fluid and the muscular action associated with mating itself. In addition to inseminating females, the

experimental males probably also performed higher levels of courtship than control males. The control males kept without females performed no courtship. When inseminated females are courted they extrude the ovipositor, which terminates the male courtship¹³. In nature, nutritional effects may be increased by food shortage, and there may be costs of sexual activity in addition to those detected here. Searching for mates¹⁴ and fighting with other males¹⁵ may be costly physiologically and these activities, together with courtship and mating, may make males more vulnerable to predation¹⁶.

Sexual activity could affect longevity in two ways. First, it may have an effect on the probability of death occurring in the next short period of time. Cessation of sexual activity at any age would then leave a fly with a life expectancy comparable with that of controls of the same age. High temperatures can have such an effect on longevity in *Drosophila subobscura*¹⁷. Second, sexual activity may have some cumulative, possibly irreversible, effect. Williams¹⁸ has suggested that senescence may be caused by the deleterious pleiotropic effects later in life of genes which have beneficial effects early in life. A deleterious long-term effect of sexual activity earlier in life could produce such a pleiotropic effect. Phenotypic correlations of the kind found in these experiments need not necessarily mean that a genetically caused change in the level of male sexual activity would alter longevity. A negative additive genetic correlation would be needed to demonstrate this and should be the subject of further experiments.

We thank Peter Calow, Brian Charlesworth, Vernon French, Paul Harvey and John Maynard Smith for helpful comments.

Received 20 August; accepted 2 November 1981.

- Gadgil, M. & Bossert, W. H. *Am. Nat.* **104**, 1 (1970).
- Stearns, S. C. *Q. Rev. Biol.* **51**, 3 (1976).
- Horn, H. S. in *Behavioural Ecology* (eds Krebs, J. R. & Davies, N. B.) 411–429 (Blackwell, Oxford, 1978).
- Rose, M. & Charlesworth, B. *Nature* **287**, 141 (1980).
- Rose, M. & Charlesworth, B. *Genetics* (in the press).
- Maynard Smith, J. *J. exp. Biol.* **35**, 832 (1958).
- Lamb, M. J. *J. Inst. Physiol.* **10**, 487 (1964).
- Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine-Atherton, Chicago, 1972).
- Krebs, J. R. & Davies, N. B. *An Introduction to Behavioural Ecology* (Blackwell, Oxford, 1981).
- Manning, A. *Nature* **194**, 252 (1962).
- Pyle, D. W. & Gromko, M. H. *Am. Nat.* **117**, 133 (1981).
- Calow, P. *Biol. Rev.* **54**, 23 (1979).
- Bastock, M. & Manning, A. *Behaviour* **8**, 85 (1955).
- Thornhill, R. *Evolution* **34**, 519 (1980).
- Dow, M. A. & von Schilcher, F. *Nature* **254**, 511 (1975).
- Thornhill, R. *Rev. Ecol. Syst.* **12**, 355 (1981).
- Clarke, J. M. & Maynard Smith, J. *J. exp. Biol.* **38**, 679 (1961).
- Williams, G. C. *Evolution* **11**, 398 (1957).

Pulmonary vein as an ectopic focus in digitalis-induced arrhythmia

D. W. Cheung*

Department of Physiology, Monash University, Clayton, Victoria 3168, Australia

The tunica media of the pulmonary vein (PV) in many mammalian species, including man, is made up of cardiac muscle^{1–3}. Brunton and Fayrer first reported that independent pulsation of the PV occurred in cats and rabbits even after activity of the heart had ceased⁴. Electrophysiological studies have also shown spontaneous electrical activity in isolated PVs of the guinea pig and a pacemaking region has been located at the distal end of the cardiac portion of the vein adjoining the smooth muscle. The intrinsic frequency of the PV was low and activity at the PV was normally coupled to the sinus rhythm⁵. Thus, the PV behaves as a subsidiary pacemaker. In the presence of digitalis-type agents, subsidiary pacemakers such as Purkinje fibres develop oscillatory afterpotentials (OAPs) which may be large enough to reach threshold^{6,7}, and it has been suggested that OAPs at Purkinje fibres leading to spontaneous action potentials may provide the underlying mechanism for ventricular arrhythmia during digitalis intoxication⁶. In contrast, atrial and ventricular muscles are not very sensitive to the arrhythmogenic action of digitalis⁸. I have therefore now investigated whether the PV can develop OAPs and act as an ectopic focus with digitalis intoxication. By recording with intracellular microelectrodes simultaneously at the PV and right atrium, I demonstrate that OAPs and repetitive activity develop at the PV in the presence of ouabain. Propagation of these triggered action potentials at the PV into the atrium results in atrial extrasystoles.

The recording arrangements in these experiments were similar to those reported previously⁵. Using male guinea pigs (300–400 g), intracellular recordings were made with glass micropipettes filled with 3 M KCl. All recordings were made at the distal end of the cardiac PV and the dorsal surface of the intact right atrium close to the vena cava. The experiments were carried out at 32–34 °C, the preparations being superfused continuously with normal or ouabain-containing physiological solutions at a rate of 4 ml min⁻¹.

In spontaneously active preparations, the PV was driven by impulses originating from the SA node⁵. In the presence of ouabain (0.5–1.0 μ M), the sinus rate decreased and became less regular, and in all seven preparations studied, after 30–40 min OAPs developed at the PV but not in atrial muscle cells (Fig. 1). Diastolic potentials between action potentials remained flat at the atrial recording site.

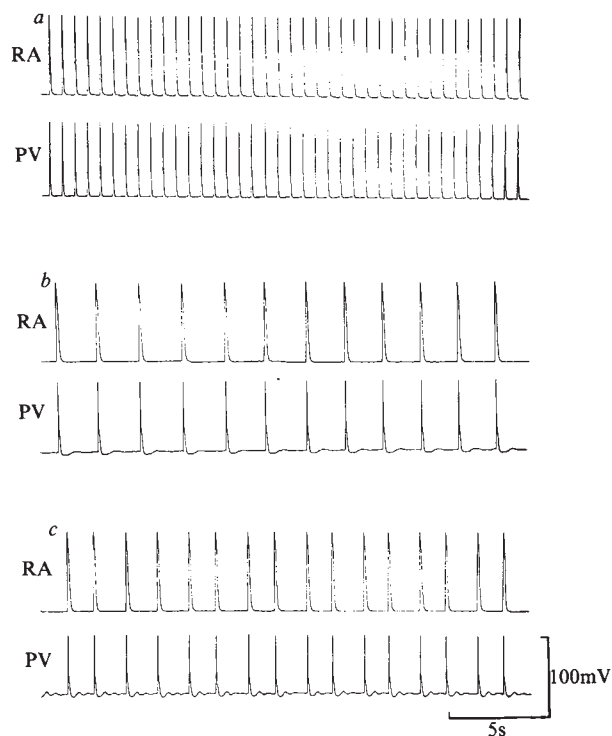


Fig. 1 Effect of ouabain (0.5 μ M) on the electrical activity of the pulmonary vein (PV) and the right atrium (RA). As demonstrated previously, action potentials at the PV are closely coupled to those at the RA⁵. After exposure to ouabain for 40 min, OAPs developed at the PV (*b*) and became very prominent at 50 min (*c*). No membrane oscillation was observed at the RA. The spontaneous frequency of the preparation decreased in the presence of ouabain.

When higher ouabain concentrations were used ($\geq 2 \mu$ M), high frequency repetitive activity appeared in six out of six preparations. Figure 2 demonstrates the onset of repetitive activity in a preparation after exposure to ouabain (2.5 μ M) for 75 min. Initially, two short bursts of extrasystoles were observed (Fig. 2*a*), soon followed by a period of repetitive activity lasting for about 2½ min at both the PV and the right atrium. Another train of repetitive activity followed after a brief pause (Fig. 2*b*).

Closer examination of the record revealed that the extrasystoles were of PV origin triggered by the sinus beats. In normal conditions, action potentials at the right atrium always preceded those at the PV⁵. The first pair of action potentials of each train of repetitive activity was also led by the atrial muscle, indicating the sinus origin of these action potentials. However, the lead was shifted to the PV in all subsequent action potentials of the train, suggesting the diversion of the pacemaking site to a region close to the PV (Fig. 2*c*). These action potentials at the PV were characterized by large diastolic depolarizations. In contrast, action potentials at the right atrium arose abruptly from the baseline membrane potential. Subthreshold OAPs were observed at the PV but not at atrial sites after termination of each train. Figure 2*d* also demonstrates the initiation of a second train of repetitive activity triggered by the sinus beat and the subsequent shift of pacemaking site to the PV.

The amplitude of action potentials at the PV decreased gradually with time in ouabain. This decrease was more marked during repetitive activity. For example, the amplitude of action potentials decreased by about 30 mV at the end of the long train in Fig. 2*b* and remained small subsequently. There was also a slight decrement in the amplitude of action potentials at the right atrium during repetitive activity.

The duration of each period of repetitive activity decreased drastically within 10 to 20 min after their appearance. For example, 5 min after the long train shown in Fig. 2*b*, only three action potentials appeared in each train (Fig. 3*a*). This further

*Present address: Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada M5S 1A8.

decreased to two action potentials after a further 2 min (Fig. 3b), and 4 min later, no extrasystoles were observed (Fig. 3c).

To demonstrate that extrasystoles originated at the PV and required a triggering input, the PV was stimulated electrically using a pair of platinum wire electrodes placed close to the recording site. It was found that repetitive activity could also be initiated with electrical stimulation (Figs 3b, 4). Figure 3b demonstrates a coupled beat in response to a single stimulus at the PV. This response was similar to those occurring spontaneously except the phase lead of both the first and second pair of action potentials were now at the PV.

Conduction block between the PV and the right atrium developed in four out of seven preparations in the presence of ouabain. Figure 4a shows an example of unidirectional block with triggered activity at the PV coupled to the sinus beats. Repetitive activity at the PV did not affect atrial activity. A more severe block is demonstrated in Fig. 4b, in which the PV was quiescent and uncoupled to atrial activity. A strong electrical pulse applied at the PV proximal to the heart elicited an action potential in the atrium and at the same time triggered repetitive activity at the PV. After severing the PV from the heart, repetitive activity could still be triggered at the PV with electrical stimulation (Fig. 4c).

I have demonstrated here that the PV can act as an ectopic focus in atrial arrhythmia with digitalis intoxication. Thus, similar to Purkinje fibres, OAPs and repetitive activity developed at the PV in the presence of ouabain, the duration of repetitive activity decreasing with time of exposure to ouabain.

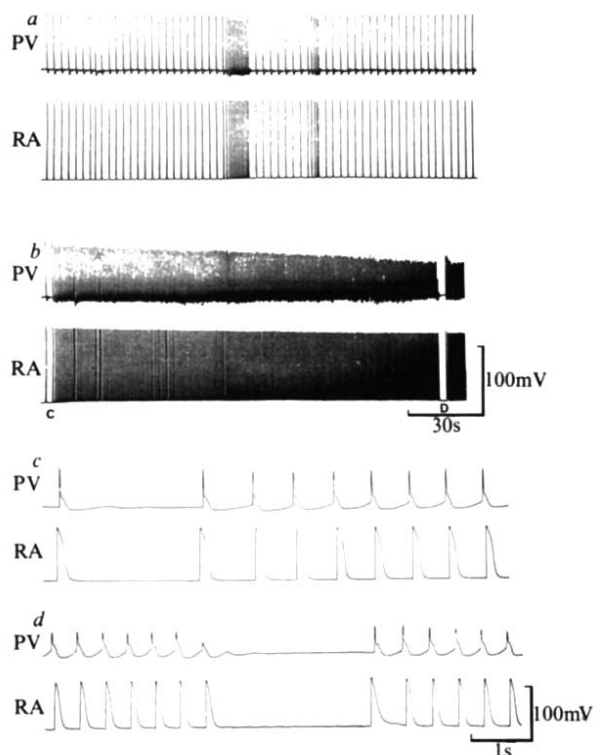


Fig. 2 Effect of a high concentration of ouabain ($2.5 \mu\text{M}$) on the spontaneous activity of an atrium-PV preparation. After 75 min in ouabain, OAPs at the PV reached threshold. Initially, only two brief bursts were observed (a), but was followed soon after by a long train of repetitive response lasting for $2\frac{1}{2}$ min (b). The first pair of action potentials of the train was characterized by the action potential at the RA preceding that of the PV (c). The phase lead then shifted to the PV in all subsequent action potentials of the train. These triggered action potentials were associated with pacemaking potentials at the PV. Subthreshold OAPs were observed at the PV following termination of each train (d). d Also shows the initiation of a second train by a sinus beat and the subsequent shift in phase lead to the PV. There was also a dramatic reduction in the amplitude of action potentials at the PV with repetitive activity (b).

The amplitude of action potentials of the PV also decreased dramatically with repetitive activity although the membrane potential did not alter significantly. A possible cause for the decrease in amplitude of action potentials and OAPs with prolonged exposure to ouabain is the accumulation of intracellular sodium. High intracellular sodium would shift the sodium

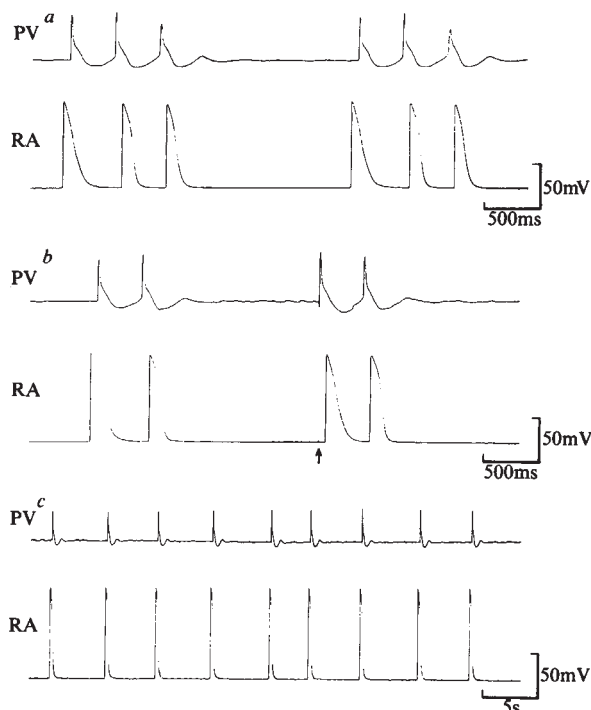


Fig. 3 The duration of repetitive activity decreased with prolonged exposure to ouabain. a Was recorded 5 min after the long train in Fig. 2b and b was recorded another 2 min later. Extrasystoles were not observed with further exposure to ouabain c. In b, an electrical stimulus (arrow) was applied close to the recording electrode at the PV to produce a coupled response similar to that occurring spontaneously.

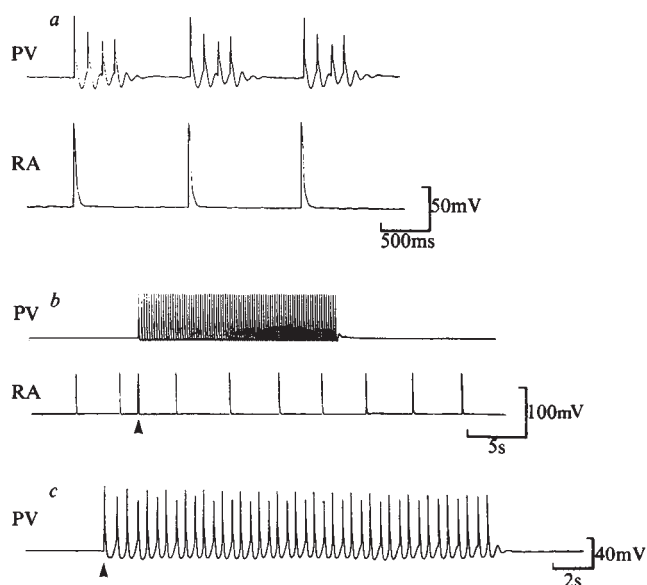


Fig. 4 Conduction block between PV and RA in the presence of ouabain ($2.5 \mu\text{M}$). a, In this preparation, repetitive activity at the PV was coupled to the sinus beats. b, The PV was not coupled to the RA and was electrically quiescent. A single electrical stimulus applied at the PV close to the heart elicited repetitive activity at the PV and an action potential at the RA. In both cases, repetitive activity at the PV did not propagate into the RA. c, After being severed from the heart, repetitive activity could still be generated at the PV with electrical stimulation.

equilibrium potential, leading to a reduction in the amplitude of action potentials. The amplitude of OAPs might also be reduced if the transient inward current responsible for OAPs was also sodium dependent⁷. The antagonistic effect of high intracellular sodium on OAP may in fact provide a mechanism for spontaneous termination of repetitive activity. With high frequency firing during repetitive activity, the amplitude of OAPs would eventually decrease to below threshold because of the additional influx of sodium with the action potentials. In the brief interval between trains of repetitive activity, the concentration of intracellular sodium may decrease, depending on the degree of poisoning of the pump, and OAPs may again reach threshold for repetitive activity.

D.W.C. is a Senior Research Fellow of the Ontario Heart Foundation.

Received 28 April; accepted 9 October 1981.

- Spach, M. S., Barr, R. C. & Jewett, P. H. *Am. J. Cardiol.* **30**, 844–854 (1972).
- De Almeida, O. A., Bohm, G. M., De Paula Carvalho, M. & De Carvalho, A. P. *J. Morph.* **135**, 409–434 (1975).
- Kramer, A. W. Jr & Marks, L. S. *J. Morph.* **117**, 135–150 (1965).
- Brunton, T. L. & Fayrer, J. *Proc. R. Soc. B25*, 174–176 (1876).
- Cheung, D. W. *J. Physiol., Lond.* **314**, 445–456 (1981).
- Ferrier, G. R., Saunders, J. H. & Mendez, C. *Circulation Res.* **32**, 600–609 (1973).
- Ferrier, G. R. *Prog. cardiovasc. Dis.* **19**, 459–474 (1977).
- Hashimoto, K. & Moe, G. K. *Circulation Res.* **32**, 618–624 (1973).
- Kass, R. S., Tsien, R. W. & Weingart, R. *J. Physiol., Lond.* **281**, 209–226 (1978).

Autoradiographic localization of GABA_B receptors in rat cerebellum

G. P. Wilkin

Department of Biochemistry, Imperial College of Science and Technology, London SW7 2BX, UK

A. L. Hudson, D. R. Hill & N. G. Bowery*

Department of Pharmacology, St Thomas's Hospital Medical School, London SE1 7EH, UK

γ -Aminobutyric acid (GABA) is a major inhibitory neurotransmitter in the mammalian cerebellum^{1–4}. Evidence for the role of GABA in this brain region stems not only from neuropharmacological studies but also from *in vitro* and *ex vivo* ligand binding studies with ³H-GABA or ³H-muscimol^{5–10}. The cerebellar concentration of GABA binding sites is higher than that elsewhere^{11,12} even though it contains less endogenous GABA than many other brain regions¹³. Using an autoradiographic technique, Kuhar and co-workers¹⁴ have recently demonstrated that these binding sites in the rat cerebellum (detected in Tris-citrate buffer) are located primarily in the granule cell layer, a finding which supports the observations of Candy and Martin¹⁵ and Olsen and Mikoshiba⁸ that ³H-GABA binding is four to six times greater in homogenates of the granule cell layer than in the molecular layer. The GABA sites detected in these studies were bicuculline sensitive and exhibited a high affinity for muscimol. We would therefore classify them as GABA_A sites because our recent observations have indicated the presence of other GABA sites (GABA_B sites) in the central nervous system which are insensitive to bicuculline^{16–18}. These sites have a high affinity for β -*p*-chlorophenyl GABA (baclofen), the GABA analogue which is devoid of activity at GABA_A sites. We describe here the autoradiographic localization of these GABA_B sites within the cerebellum and show that, unlike GABA_A sites, they are confined almost exclusively to the molecular layer.

As our previous studies on GABA_B site binding have been limited to membranes prepared from whole rat brain^{18,19}, it was first necessary to examine the regional distribution of these binding sites before studying their cellular location within any

one region. GABA_B sites can be readily detected in crude synaptic membrane preparations using either ³H-GABA (50 Ci mmol⁻¹, Amersham) or ³H-baclofen (8.8 Ci mmol⁻¹, Ciba-Geigy) as the ligand. It is essential, however, that the divalent cations Ca²⁺ or Mg²⁺ (1–5 mM) are present in the incubation medium and that Tris-HCl (50 mM, pH 7.4) is used instead of Tris-citrate buffer. Citrate presumably chelates any available divalent ions to prevent binding. High concentrations of Tris will also reduce GABA_B site binding. In their autoradiographical studies, Palacios *et al.*¹⁴ used 0.3 M Tris buffer to detect GABA binding sites. At this concentration of Tris we found that whereas the GABA_A site binding of ³H-GABA (30 nM) is unaffected, binding to GABA_B sites is 57% lower than in 50 mM Tris. Consequently, the incubation and washing solution used for slices in the present study was 50 mM Tris-HCl buffer containing 190 mM sucrose (2.5 mM CaCl₂ was added to the solution used for GABA_B site binding but was absent in GABA_A site experiments). This isotonic solution (270 mosmol l⁻¹) did not depress GABA_B site binding and maintained the morphological integrity of the tissue slices for autoradiography.

For the initial distribution studies, separated brain regions were homogenized in 15 volumes of ice-cold sucrose (0.32 M) and pellets obtained as indicated in Table 1 legend, stored at –20 °C and used for assay within 3 days. The pellets were then thawed, washed four times and incubated with fixed concentrations of ³H-GABA (10 nM) or ³H-baclofen (20 nM). These are equivalent concentrations of active ligand, as ³H-baclofen is a racemate and the (–)-isomer is >100 times more potent than the (+)-isomer as an agonist at GABA_B sites.

Binding to GABA_B sites was evident in all the brain regions studied and there was a striking correlation between the

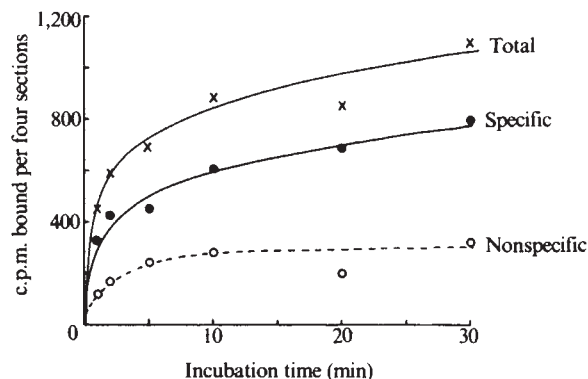


Fig. 1 Time course of ³H-GABA binding to GABA_B sites in cerebellar slices. Serial sections of a cerebellum were obtained from a male CSE rat anaesthetized with pentobarbitone before intracardiac perfusion with 0.1% paraformaldehyde in 0.01 M phosphate buffered saline (pH 7.4). The cerebellum was hemisected and immersed in isopentane at –40 °C. Sagittal sections (10 μ m) were mounted on glass slides (four per slide) and allowed to dry at ambient temperature for 1 h before storage at –15 °C for at least 16 h. For the binding assay slides were removed from storage for 45 min before immersion in 250 ml Tris-HCl (50 mM, pH 7.4) containing CaCl₂ (2.5 mM) and sucrose (190 mM) for a further 50 min. Each slide was then air dried for 10–15 min after the removal of any excess fluid. Incubation solution (100 μ l) containing ³H-GABA (50 nM) and isoguvacine (40 μ M, to suppress Ca²⁺-independent binding to GABA_A sites) with (○) or without (×) 100 μ M ($\pm$) baclofen was applied to each slide and left for 1, 2, 5, 10, 20 or 30 min. The slides were then shaken separately to remove excess incubation solution and rinsed twice very briefly in fresh radioactive-free solution. After drying, the slides were cut and placed in scintillation vials with 0.4 ml distilled water. Scintillation fluid (10 ml Packard ES 299) was added to each vial 30 min later. The c.p.m. values are means of three determinations at each incubation time. Nonspecific values were those obtained in the presence of unlabelled baclofen. The specific values (●) are the differences between total and nonspecific binding values determined at each incubation time.

* To whom correspondence should be addressed.

Table 1 GABA receptor distribution in rat brain

Brain region	B site		A site
	³ H-baclofen (20 nM) (fmol per mg protein)	³ H-GABA (10 nM) (fmol per mg protein)	³ H-GABA (10 nM) (fmol per mg protein)
Cerebellum	76.1 ± 12.5	73.9 ± 5.9	331 ± 45
Cerebral cortex	22.7 ± 5.5	28.7 ± 2.1	66.5 ± 6.9
Brain stem	13.9 ± 4.3	13.1 ± 2.4	39.8 ± 7.6
Spinal cord	11.4 ± 2.6	10.7 ± 0.8	31.8 ± 6.8
Mid-brain + forebrain	23.4 ± 4.0	18.4 ± 4.6	51.2 ± 9.4

The regional distribution of GABA binding sites in rat brain membrane preparations. In each experiment separated brain regions from two rats were homogenized in 15 vol ice-cold sucrose (0.32 M) using a Potter-Elvehjem homogenizer fitted with a Teflon pestle. The homogenates were centrifuged at 1,000g for 10 min and the supernatants re-centrifuged at 20,000g for 20 min. The pellets were then resuspended in the original volume of distilled water before centrifuging at 48,000g for 20 min. The final pellets were frozen at -15 °C for at least 16 h. For the assay each pellet was allowed to thaw at room temperature for 30–40 min before resuspending and washing four times in 6 ml Tris-HCl buffer (pH 7.4, 50 mM) containing CaCl₂ (and isoguvacine 40 μM to suppress binding to GABA_A sites) for GABA_B site or Tris-citrate buffer (pH 7.4, 50 mM) for GABA_A site binding. The final pellet was resuspended in either medium (1–1.5 mg protein per 0.8 ml solution determined by the method of Lowry *et al.*²²) for the assay. To each 0.8-ml aliquot 0.1 ml of incubation fluid with or without 1 mM (±) baclofen (GABA_B site) or 1 mM isoguvacine (GABA_A sites) was added. A further 0.1 ml containing ³H-baclofen or ³H-GABA was added to provide final concentrations of 20 nM or 10 nM, respectively. Each mixture was incubated for 10 min at 20 °C and then centrifuged at 7,000g for 10 min. The supernatant was aspirated off and the pellet blotted dry before the addition of 100 μl Soluene and subsequent scintillation fluid. The data are from six separate experiments in each of which binding to GABA_A and GABA_B sites was determined simultaneously. Values are mean ± s.e.

amounts of ³H-GABA and ³H-baclofen specifically bound in each region. The concentration of GABA_A sites was clearly much greater than that of GABA_B sites in all regions, with an overall ratio of GABA_A:GABA_B sites of 3 or 4:1. The most important finding was the high concentration of GABA_B sites in the cerebellum, which prompted us to examine the binding in this brain region more closely using an autoradiographic technique.

Brains from rats anaesthetized with pentobarbitone sodium (60 mg per kg intraperitoneally) were fixed by intracardiac perfusion with 0.1% paraformaldehyde in phosphate-buffered saline (0.01 M pH 7.4, 10 min perfusion with 200 ml). In some experiments higher fixative concentrations were used to see whether this affected the subsequent binding. At concentrations below 0.5% paraformaldehyde saturable binding was readily detected and apparently independent of the fixative concentration whereas above this concentration (up to 4%) it was markedly reduced and extremely variable. Sagittal slices (10 μm thick) of cerebellum were obtained by cryostat sectioning and four slices placed on single glass slides. The slices were allowed to dry at ambient temperature for 1 h before freezing at -20 °C. The tissue was stored in this manner for up to 1 week without any noticeable impairment in its ability to bind ³H-GABA to either GABA_A or GABA_B sites.

³H-GABA (50 Ci mmol⁻¹) rather than ³H-baclofen was used for the experiments because it has a higher specific activity than ³H-baclofen and no inactive enantiomer. Qualitatively similar results were obtained with ³H-baclofen binding to GABA_B sites but, as expected, the background radioactivity was higher and the total radiolabelling lower than with ³H-GABA.

For the binding experiments the slices were allowed to thaw and dry for 45 min and then washed for 50 min by immersion in incubation solution (250 ml per 30 slides). After drying, 0.1 ml incubation solution containing 40 nM ³H-GABA with or

without unlabelled displacing agents was applied to each slide. This was subsequently removed by shaking and two brief rinses in large volumes of the appropriate Tris-sucrose solution. The slices were then either placed in a vial for scintillation spectrometry or kept for autoradiography (see Fig. 3 legend).

Figure 1 shows the time course of ³H-GABA binding to GABA_B sites in these conditions. The sections were incubated with ³H-GABA in the presence of 40 μM isoguvacine (to prevent GABA_A site binding) with or without unlabelled (±)baclofen (100 μM). The total amount of ³H-GABA bound increased with time for at least 20 min, whereas in the presence of unlabelled baclofen the amount bound (nonspecific binding) was constant after 5 min. The specific portion of binding (total minus nonspecific) remained constant after 20 min incubation and >85% of this had occurred by 15 min. All subsequent studies were performed using incubations of 15 min, by which time the specific portion comprised 75.4 ± 2.3% (mean ± s.e.m., *n* = 9; triplicate determinations) of total binding. In the same conditions ³H-GABA binding to GABA_A sites formed 90.1 ± 1.6% (*n* = 4 in triplicate) of the total binding.

Scatchard analysis of specific ³H-GABA binding to GABA_B sites in cerebellar slices indicated the presence of a single saturable binding component *K*_d = 159 nM, *B*_{max} = 83.6 fmol per 4 slices (Hill coefficient = 0.97). The apparent affinity of this component was therefore slightly lower in slices than previously observed in homogenates (*K*_d = 77 nM), perhaps reflecting the presence of endogenous inhibitors in the intact slices which may be removed in the homogenate preparations.

³H-GABA which was specifically bound to GABA_B sites in slices could be displaced by analogues shown previously to be

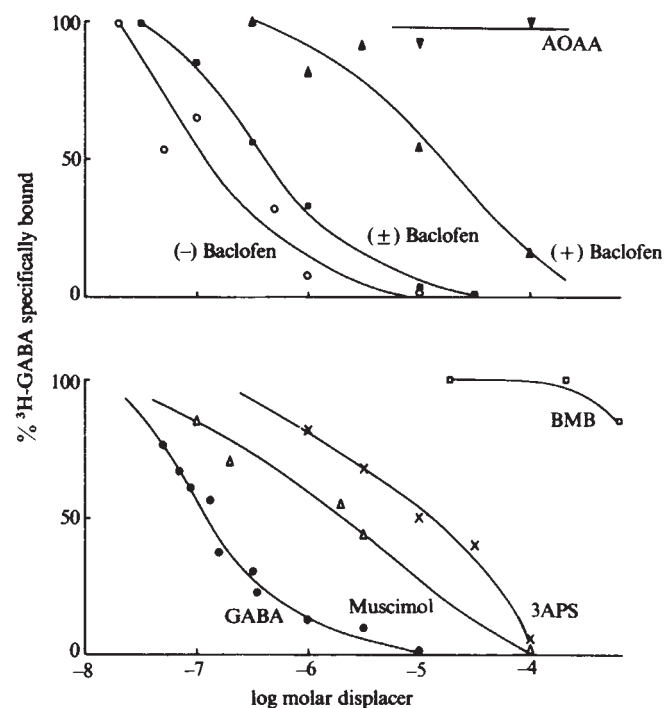


Fig. 2 Displacement of specifically bound ³H-GABA from GABA_B sites in rat cerebellar slices by unlabelled GABA-related substances. Cerebellar slices were obtained and treated in the same manner as described in Fig. 1 legend except that all slices were incubated for 15 min and various concentrations of different unlabelled analogues were added to the incubation solution. Total specific binding in any experiment was that portion of ³H-GABA displaced by 100 μM (±)baclofen. The displacement produced by each concentration of analogue is expressed as a percentage of that occurring with (±)baclofen (100 μM). The data have been separated for clarity. Each value is the mean of three determinations (four slices per determination). Standard errors were <5%, ○, (-)Baclofen; ■, (±)baclofen; ▲, (+)baclofen; ▼, aminoxyacetic acid; ●, GABA; △, muscimol; ×, 3-amino-propanesulphonic acid; □, bicuculline methobromide.

active at GABA_B sites in homogenates as well as whole tissue assays. Slices were incubated simultaneously with ³H-GABA and various concentrations of unlabelled displacer (as well as 40 μ M isoguvacine) and the combined data from these experiments are shown in Fig. 2. GABA and (–)baclofen were equipotent in displacing ³H-GABA, with IC₅₀ values of 120 nM. The IC₅₀ values for the other analogues were 400 nM (±)baclofen, 2 μ M for muscimol, 12.6 μ M for 3-amino-propanesulphonic acid and 15.8 μ M for (+)baclofen, an order of potency which accords with the analogue specificity observed previously in homogenates¹⁶. As expected, bicuculline metho-

bromide (100 μ M) and the GABA transaminase inhibitor aminoxyacetic acid (AOAA) were inactive. In the latter case, as both GABA and AOAA would be expected to bind to the active site of the enzyme, the absence of effect of AOAA suggests that the transaminase enzyme is unlikely to be the binding site in slices.

Inhibitors of GABA transport—*cis*-3-aminocyclohexane carboxylic acid (ACHC), nipecotic acid and 2,4 diaminobutyric acid (DABA)—produced no displacement at 10 μ M and less than 20% displacement at 100 μ M; the IC₅₀ values were all $\gg$ 1 mM. It is therefore unlikely that the binding site is a trans-

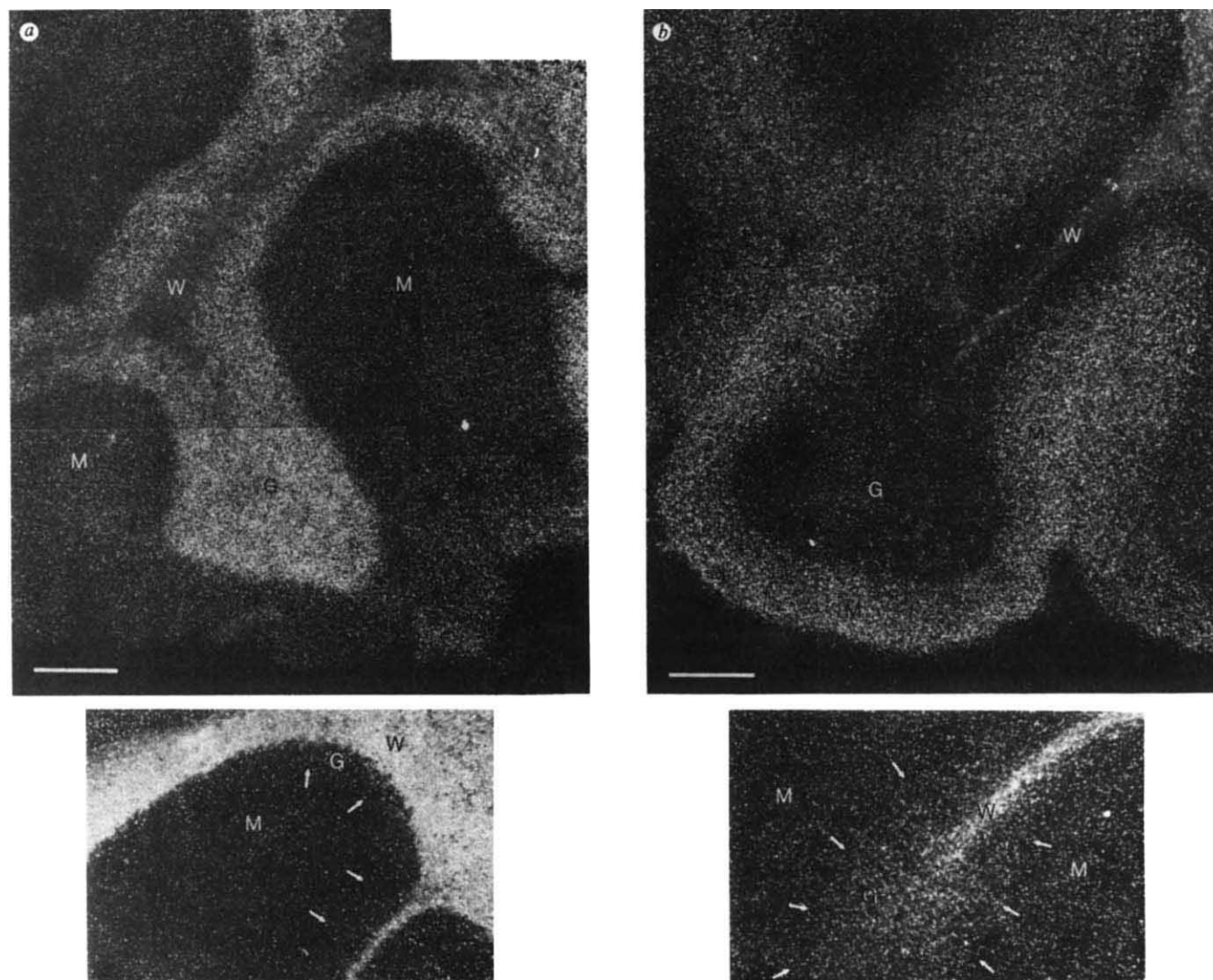


Fig. 3 Light microscopic darkfield autoradiograms of ³H-GABA binding to GABA_A sites (a) and GABA_B sites (b) in rat cerebellum. Cerebellar cryostat sections were obtained as described in Fig. 1 legend. After thawing and drying for 45 min the sections were immersed in Tris-HCl solution containing sucrose (190 mM) with (GABA_B sites) or without (GABA_A sites) CaCl₂ (2.5 mM) for 50 min. For GABA_A site binding, 100 μ l of incubation fluid containing ³H-GABA (50 nM) with (nonspecific binding) or without (specific binding) unlabelled isoguvacine (100 μ M) was applied for 15 min. This was removed by shaking and two brief rinses in large volumes of incubation medium. For GABA_B site binding the same procedure was followed except that the incubation medium contained CaCl₂ (2.5 mM) and isoguvacine (40 μ M) to suppress binding to GABA_A sites. Nonspecific binding was determined by the addition of (±)baclofen (100 μ M). After drying, the slides were prepared for autoradiography following the procedure of Young and Kuhar²³. Coverslips that had been dipped in Ilford K5 emulsion (emulsion: water 1:1.5) were attached to the slides with Loctite glue and batches of six slides were held together with bulldog clips and exposed for up to 12 days at +4 °C in a sealed box containing silica gel. Slides were developed at 16 °C in Kodak 19b developer (3.5 min) and fixed in 25% sodium thiosulphate solution. Following several washes in glass distilled water the slides were dried overnight and the coverslips were then permanently fixed in position with DPX mounting medium. Sections were viewed under dark field optics with a Leitz Ortholux II microscope and photographs taken with Ilford FP4 film. GABA_A sites (a): The greatest concentration of silver grains can be seen over the granule cell layer (G) and a lesser concentration over the molecular layer (M). Very few grains are present over the white matter tract (W). The smaller photograph below shows nonspecific binding. The boundary between the molecular and granule cell layers is indicated by the arrows. GABA_B sites (b): In contrast to a, the greatest accumulation of silver grains is over the white matter tract (W) with a lesser amount over the granule cell layer (G) and white matter (W). In the section below, displaying nonspecific binding, the boundary of the molecular and granule cell layer is again delineated by arrows. The white matter tracts in such sections show varying degrees of light scattering; here it is greater in the blank (lower picture) than in the experimental section. Such light scattering, however, does not hinder the counting of silver grains. Scale bar, 300 μ m.

Table 2 Autoradiographic grain densities over cerebellar layers following ^3H -GABA binding to GABA_B, GABA_A and both sites together in 10- μm cryostat sections

Layer	GABA _B sites			GABA _A sites			GABA _B and GABA _A sites		
	Total	Nonspecific	Total – nonspecific	Total	Nonspecific	Total – nonspecific	Total	Nonspecific	Total – nonspecific
White matter	1.21 ± 0.16	1.08 ± 0.07	0.13	1.44 ± 0.12	0.87 ± 0.08	0.57	1.24 ± 0.36	0.76 ± 0.06	0.48
Granule cell layer	1.64 ± 0.17	1.12 ± 0.08	0.52	9.88 ± 0.62	0.94 ± 0.08	8.94	10.47 ± 0.83	0.85 ± 0.05	9.62
Molecular layer	3.82 ± 0.17	1.09 ± 0.05	2.73	3.85 ± 0.18	1.10 ± 0.19	2.75	6.38 ± 1.06	0.95 ± 0.05	5.43

Densities were quantified after 1 week's exposure using an eyepiece graticule. In each case the mean $\pm$ s.e.m. was obtained after counting four separate areas each of 5,400 μm^2 on two slides. Values are grains per 100 μm^2 .

port recognition site for GABA, and this is supported by the Na⁺ independence of binding in slices.

The close comparison between the pharmacological specificity observed in slices and homogenates indicated that it was appropriate to proceed with autoradiography in the slices. Thus, slides which had been incubated in ^3H -GABA to label GABA_A or GABA_B sites were dried, coverslips coated with emulsion (Ilford K5) were placed in contact with the sections, and the slides were stored in boxes at +4 °C for 6–12 days before developing. Figure 3 gives results from two slices from the same cerebellum after 12 days' exposure. Figure 3a shows the distribution of silver grains after incubation with ^3H -GABA (50 nM) in the absence of Ca²⁺ and presence of 100 μM ($\pm$)baclofen; the associated background picture was obtained by the further addition of isoguvacine (100 μM). Figure 3b shows the distribution after incubation with ^3H -GABA in the presence of 2.5 mM CaCl₂ and 40 μM isoguvacine; ($\pm$)baclofen (100 μM) was further added to obtain the background picture. The distribution of silver grains in the two incubation conditions was markedly different. GABA_A site binding (Fig. 3a) occurred primarily in the granule cell layer whereas GABA_B site binding (Fig. 3b) was confined to the molecular layer. The grain counts for these individual layers are shown in Table 2.

Concomitant experiments using slices from the same animal were subjected to analysis by scintillation spectrometry. The mean radioactivity (c.p.m. values per four slices, $\pm$ s.e.m., $n = 6$) for GABA_A and GABA_B sites, respectively, were 3,591 $\pm$ 269 (background 393 $\pm$ 156) and 962 $\pm$ 3 (background 236 $\pm$ 66). The specific portions were therefore 3,198 and 726 c.p.m., respectively, which represents a ratio of 4.4 : 1. When slices were incubated in ^3H -GABA in the presence of Ca²⁺ alone to obtain the total GABA_A and GABA_B site binding, the value was 4,203 $\pm$ 303 c.p.m. (background 236 $\pm$ 34). The specific portion was then 3,967 c.p.m., which was identical with the sum of the separate data (3,198 $\pm$ 726 = 3,924 c.p.m.).

The present study has shown that Ca²⁺-dependent GABA binding sites (GABA_B sites) in the rat cerebellum are localized almost entirely within the molecular layer. Whether these sites are functional remains to be seen, although it may be pertinent that whereas the number of presumed GABA nerve terminals is similar in the molecular and granule cell layer, 'classical' GABA_A sites are unequally distributed¹⁴. The presence of GABA_B sites may explain this apparent deficit of GABA receptors in the molecular layer.

We do not know the cell type(s) on which the GABA_B sites are located. Evidence from release studies has suggested that they are present on presynaptic terminals where they modulate transmitter release^{16,20}. However, we know of no morphological evidence for axo-axonic synapses on terminal boutons in the molecular layer²¹. Thus, allowing for the fact that such a population may not have been recognized, other possibilities are that GABA_B sites are (1) non-innervated (2) present postsynaptically on morphologically 'conventional' axo-dendritic synapses or (3) present on glial cells. This third possibility seems unlikely because there is no autoradiographic evidence of binding over fibrous astrocytes present in white matter, and pellets enriched in synaptic membranes show a high degree of binding. Work is in progress to investigate these possibilities.

D.R.H. is an SRC scholar. We thank St Thomas's Hospital Research Endowment Fund for financial support, Ciba-Geigy

Ltd for supplies of ^3H -baclofen and the isomers of baclofen, Drs P. Krogsgaard-Larsen and J. Collins for isoguvacine and (+)bicuculline methobromide, and Mrs J. M. Andrews for typing the manuscript.

Received 8 July; accepted 14 October 1981.

- Höfkelt, T. & Ljungdahl, A. *Expl Brain Res.* **14**, 354–362 (1972).
- Curtis, D. R. & Johnston, G. A. R. *Ergebn. Physiol.* **69**, 97–188 (1974).
- Ribak, C. E., Vaughn, J. E. & Saito, K. *Brain Res.* **143**, 315–332 (1978).
- Wilkin, G. P. *et al. Brain Res.* **216**, 11–33 (1981).
- Chan-Palay, V. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1024–1028 (1978).
- Chan-Palay, V. & Palay, S. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2977–2980 (1978).
- Hörng, J. S. & Wong, D. T. *J. Neurochem.* **32**, 1379–1386 (1979).
- Olsen, R. W. & Mikoshiba, K. *J. Neurochem.* **30**, 1633–1636 (1978).
- Simantov, R., Oster-Granite, M. L., Herndon, R. L. & Snyder, S. H. *Brain Res.* **105**, 365–371 (1976).
- Kingsbury, A. E., Wilkin, G. P., Patel, A. J. & Balazs, R. *J. Neurochem.* **35**, 739–742 (1980).
- Enna, S. J., Beaumont, K. & Yamamura, H. I. in *Amino Acids as Chemical Transmitters* (ed. Fonnum, F.) 487–492 (Plenum, New York, 1978).
- Lloyd, K. G. & Dreksler, S. in *Amino Acids as Chemical Transmitters* (ed. Fonnum, F.) 457–466 (Plenum, New York, 1978).
- Fahn, S. & Côté, L. J. *J. Neurochem.* **15**, 209–213 (1968).
- Palacios, J. M., Young, W. S. & Kuhar, M. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 670–674 (1980).
- Candy, J. C. & Martin, I. L. in *Proc. Colloq. Receptors, Neurotransmitters and Peptide Hormones*, Capri, 123 (1979).
- Bowery, N. G. *et al. Nature* **283**, 92–94 (1980).
- Bowery, N. G., Doble, A., Hill, D. R., Hudson, A. L., Shaw, J. & Turnbull, M. J. *Brain Res. Bull.* **5**, Suppl. 2, 497–502 (1980).
- Hill, D. R. & Bowery, N. G. *Nature* **290**, 149–152 (1981).
- Bowery, N. G., Hill, D. R. & Hudson, A. L. *Br. J. Pharmac.* **74**, 222P–223P (1981).
- Bowery, N. G. *et al. Eur. J. Pharmac.* **71**, 53–70 (1981).
- Palay, S. L. & Chan-Palay, V. *Cerebellar Cortex, Cytology and Organisation* (Springer, Heidelberg, 1974).
- Lowry, D. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).
- Young, W. S. & Kuhar, M. J. *Brain Res.* **179**, 255–270 (1979).

Neurotensin receptors are located on dopamine-containing neurones in rat midbrain

Jose M. Palacios & Michael J. Kuhar*

Departments of Neuroscience, Pharmacology and Experimental Therapeutics and Psychiatry and the Behavioral Sciences, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Neurotensin is a putative peptide neurotransmitter in the mammalian brain; neurotensin-containing neurones and neurotensin receptors have been identified in the brain both biochemically and histochemically^{1–7}. There is strong evidence for an interaction between dopaminergic systems and neurotensin^{7–11}. Related to this, light microscopic radiohistochemical studies have shown a dense localization of neurotensin receptors in the substantia nigra zona compacta and the adjacent ventral tegmental area of the rat midbrain⁷. These studies suggest a high concentration of neurotensin receptors on dopaminergic cell bodies. Here we report a large, local depletion of neurotensin receptors in the substantia nigra zona compacta after a local 6-hydroxydopamine injection, indicating that endogenous neurotensin or related drugs could have potent effects on dopaminergic function in the brain.

* To whom correspondence could be sent.

Male, Sprague-Dawley rats (225–275 g) were used in our experiments. 6-Hydroxydopamine (Sigma) was dissolved in saline containing 0.2 mg ml^{-1} ascorbate to give a final concentration of 2 mg ml^{-1} and $6 \mu\text{g}$ of drug were injected over a 5 min period with a Hamilton $5 \mu\text{l}$ syringe, which was placed stereotactically over the substantia nigra as previously described¹². Three animals were injected with drug and three were injected with saline as controls. Four weeks after injection, the animals were killed and tissues prepared for receptor autoradiography as previously described^{7,13,14}. Sections ($8 \mu\text{m}$) of midbrain were thaw-mounted on to microscope slides and neurotensin receptors labelled by incubating the sections in $4 \text{ nM } ^3\text{H}$ -neurotensin ($64.3 \text{ Ci mmol}^{-1}$; New England Corporation) in 0.17 M Tris-HCl buffer $\text{pH } 7.6$ at ice-bath temperatures. The buffer contained 0.05% boiled bovine serum albumin (BSA) and 0.1 ml bacitracin. After incubation, the sections were washed twice for 5 min in the same buffer at 2°C to reduce nonspecific binding. Blank values were obtained by adding $5 \mu\text{M}$ neurotensin to parallel incubations. Extensive biochemical studies have shown that binding of neurotensin to the slide-mounted tissue sections in these conditions is predominantly to the neurotensin receptor⁷. Autoradiographs, produced by the apposition of ^3H -Ultrofilm (LKB) in X-ray film cassettes¹⁴, were exposed for 30 days at 2°C . After development, the autoradiographic grain densities were determined quantitatively using a micro-

Table 1 6-Hydroxydopamine injection in the rat substantia nigra affects neurotensin receptor binding

Brain area	Neurotensin receptor binding (fmol per mg protein)		
	Control side	Lesion side	% Change
Substantia nigra (pars compacta)	128.3 ± 25.8	12.3 ± 7.9	-90 ($P < 0.01$)
Ventral tegmental area	197.7 ± 25.9	226.3 ± 21.7	$+15^*$
Rhinal sulcus	175.2 ± 13.3	177.0 ± 20.4	$+1^*$
Cortical nucleus of the amygdala	161.1 ± 19.0	162.3 ± 28.3	$+1^*$

Data are mean $\pm$ s.e.m. of at least two tissue sections from each of three different animals. Data were obtained by measuring the optical densities of the autoradiograms as described in the text. By using autoradiographic standards, the optical densities were converted to fmol of ligand bound. Blank values of receptor binding were quite low and have been subtracted. Thus, the data represent specific receptor binding.

* No significant difference, Student's *t*-test.

densitometer (Gamma Scientific) with a $100\text{-}\mu\text{m}$ aperture. Autoradiographic standards were included in each experiment so that grain densities could be transformed to fmol of neurotensin bound.

In these rat midbrain sections, the receptor distribution was identical to that previously reported⁷. High densities were found in the substantia nigra as well as in the adjacent ventral tegmental area. Neuronal elements containing immunoreactive neurotensin have been detected in this area³. High densities of receptors were also found in the cortical nucleus of the amygdala, rhinal sulcus and in parts of the cerebral cortex. Serial, adjacent sections incubated with excess neurotensin to provide blanks, did not produce any significant elevated grain densities in these areas.

Injection of 6-hydroxydopamine caused a reproducible and substantial loss of cell bodies in the substantia nigra zona compacta. There was also a striking loss of neurotensin receptors in the substantia nigra that did not extend into the medial ventral tegmental areas (Fig. 1). No loss of receptor occurred in the saline-injected controls (data not shown). Thus, there was receptor loss only in areas showing cell destruction. Biochemical studies indicated a 75–80% reduction in dopamine levels in the ipsilateral striatum confirming a loss of dopamine cells in the injected substantia nigra (data not shown). There was no nonspecific loss of cells at the injection sites as benzodiazepine receptor levels were not decreased substantially in adjacent serial sections (data not shown).

To quantitate the changes in receptor distributions, the autoradiographic grain density was determined by microdensitometry (Table 1). There was a 90% loss of neurotensin receptors on the injected side of the substantia nigra. By contrast, areas unaffected by the lesion, such as the ventral tegmental area, rhinal sulcus and cortical nucleus of the amygdala, did not show any significant change in receptor density.

These results indicate that at least some dopamine-containing neurones in the ventral midbrain of the rat have very high densities of neurotensin receptors. This is corroborated by the recent finding of Andrade and Aghajanian¹⁵ that neurotensin is a highly potent and reliable excitant of dopaminergic neurones in iontophoretic studies. Our results, taken together with those of others^{7–11,15}, agree with the general idea of a neurotensin-dopamine interaction in brain, probably at several sites^{3,7–11}. Dopaminergic neurones are presumably affected by other peptides including enkephalins¹⁶, cholecystokinin¹⁷ and substance P¹⁸. Thus, the development of peptide agonists and antagonists may be useful in the treatment of syndromes involving dopaminergic neurones.

We thank Ms Darlene Weimer for clerical assistance, Mrs Naomi Taylor, Ms Anne Foard and Mrs Theresa Kopajtic for technical assistance and Dr Joseph Coyle and Mr Robert Zaczek for the measurement of the striatal dopamine levels. This work

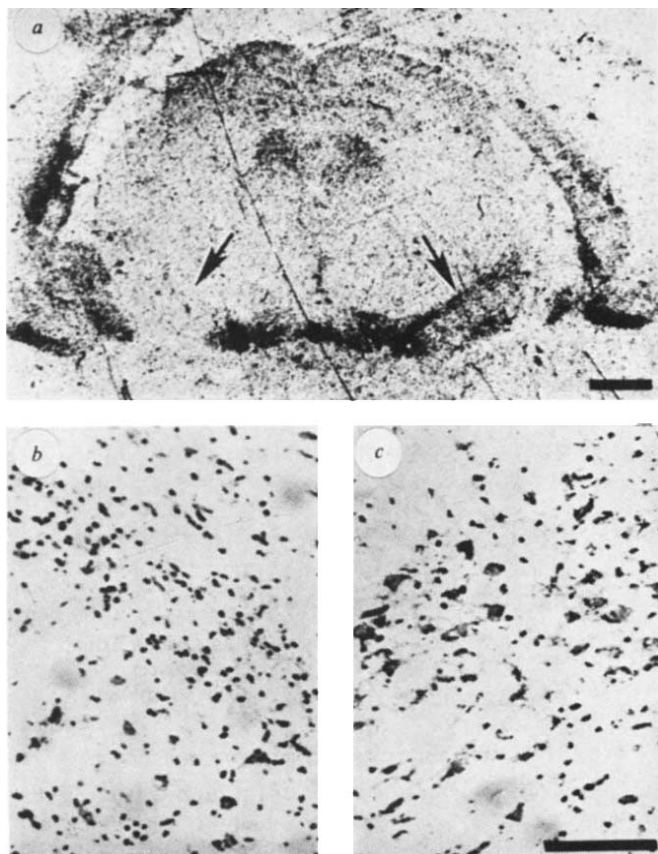


Fig. 1 Autoradiographic localization of neurotensin receptors to dopaminergic neurones in rat substantia nigra zona compacta. Injection of 6-hydroxydopamine results in a loss of neurotensin receptors and dopamine-containing cells in the substantia nigra zona compacta (see text for details). *a*, Photomicrograph of the autoradiographic image of the rat midbrain on ^3H -Ultrofilm. Note the loss of radioactive ligand binding on the left (at arrow) where the 6-hydroxydopamine was injected, compared with the binding on the right where receptor densities parallel the distribution of dopamine-containing cells. Scale bar, 1 mm . *b*, *c*, Higher power photomicrographs of the cresyl violet-stained section used to produce the image in *a*. The photomicrographs were taken at the arrows in *a*. Note the loss of cell bodies in the zona compacta in *b* (corresponding to the injected left side in *a*) compared with the opposite uninjected side, shown in *c*. Scale bar, $100 \mu\text{m}$.

was supported by USPHS grants MH25951 and MH00053 (M.J.K.) and TW02583 (J.M.P.).

Received 15 June; accepted 12 October 1981.

- Kitabgi, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **74**, 1846–1850 (1977).
- Lazarus, L., Brown, M. & Perrin, M. *Neuropharmacology* **16**, 625–629 (1977).
- Uhl, G. R. & Snyder, S. H. in *The Role of Peptides in Neuronal Function* (eds Barker, J. L. & Smith, T. G. Jr) 509–543 (Dekker, New York, 1980).
- Uhl, G. R., Kuhar, M. J. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4059–4063 (1977).
- Uhl, G. R., Bennett, J. P. Jr & Snyder, S. H. *Brain Res.* **130**, 299–313 (1977).
- Uhl, G. R. & Snyder, S. H. *Eur. J. Pharmac.* **41**, 89–91 (1977).
- Young, W. S. III & Kuhar, M. J. *Eur. J. Pharmac.* **59**, 161–163 (1979); *Brain Res.* **206**, 273–285 (1981).
- Nemeroff, C. B. *et al.* *Brain Res.* **195**, 69–84 (1980).
- Nemeroff, C. B. *Biol. Psychiat.* **15**, 283–302 (1980).
- Ervin, G. N., Birkemo, L. S., Nemeroff, C. B. & Prange, A. J. Jr, *Nature* **291**, 73–76 (1981).
- Govoni, S., Hong, J. S., Yang, H.-Y. T. & Costa, E. *J. Pharmac. exp. Ther.* **215**, 413–417 (1980).
- Murrin, L. C., Gale, K. & Kuhar, M. J. *Eur. J. Pharmac.* **60**, 229–235 (1979).
- Young, W. S. III & Kuhar, M. J. *Brain Res.* **179**, 255–270 (1979).
- Palacios, J. M., Niehoff, D. L. & Kuhar, M. J. *Neurosci. Lett.* **25**, 101–105 (1981).
- Andrade, R. & Aghajanian, G. K. *Soc. Neurosci. Abstr.* **7**, 573 (1981).
- Pollard, H., Llorens, C., Schartz, J. C., Gros, C. & Fray, F. *Brain Res.* **151**, 392–398 (1978).
- Hokfelt, T. *et al.* *Nature* **285**, 476–478 (1980).
- Michelot, R., Levie, V., Guirguieff-Chesselet, M. E., Cheramy, A. & Glowinski, J. *Life Sci.* **24**, 715–724 (1979).

Isolation and biogenesis of a new peptide from pancreatic islets

Thue W. Schwartz* & Howard S. Tager

Department of Biochemistry, University of Chicago, Chicago, Illinois 60637, USA

Pancreatic polypeptide (PP) was discovered as a major peptide contaminant of insulin preparations^{1,2}. The peptide is stored in a distinct endocrine cell type³ that is the predominating islet cell in the duodenal part of the pancreas⁴ and it is released in response to physiological events, including food intake⁵ and decrease in blood glucose^{6,7}. Furthermore, infusions of PP mimicking physiological fluctuations in plasma concentrations of the peptide affect gastrointestinal functions, by inhibiting exocrine pancreatic secretion^{8–10} and relaxing the gallbladder^{10,11}. Although the PP cell is more sensitive to small reductions in blood glucose than is the glucagon cell⁶, many studies have failed to demonstrate any metabolic effect of PP^{12,13}. It is recognized that the biogenesis of secretory peptides by conversion of larger precursors results in multiple peptide fragments of which more than one might be bioactive^{14–16}. Recently we identified a precursor of PP with a molecular weight (M_r) of 8,000–10,000 and showed, by pulse-chase experiments and peptide mapping, that this biosynthetically labelled precursor was converted into a low molecular weight peptide in addition to PP¹⁷. We report here the isolation and primary structure of this icosapeptide and show how it arises through an intermediate form by post-translational modification of the COOH-terminal part of the PP precursor. This second product of pro-pancreatic polypeptide conversion might fulfill the physiological potential of the major endocrine cell type of the duodenal pancreas.

To obtain amounts of the co-synthesized pro-PP-related product sufficient for structural, immunological, and biological studies, we used the biosynthetically labelled peptide as a probe. Duodenal lobes of canine pancreas which are especially rich in PP¹⁸ were extracted in acid ethanol and gel-filtered. As shown in Fig. 1a, a peak of absorbance (280 nm) eluted at the position of the biosynthetically labelled fragment of the PP precursor (fractions 125–135). This peak was not obvious when extracts were prepared from the splenic part of the pancreas, which is deficient in PP (data not shown). Only one major peptide eluting with the labelled fragment of the PP precursor was detected by polyacrylamide gel electrophoresis (Fig. 1a inset). (The proinsulin connecting peptide (C-peptide) which elutes in a similar position neither absorbs at 280 nm nor is it fixed and stained in the polyacrylamide gel system used.) The peptide was further

purified by DEAE-anion-exchange chromatography (Fig. 1b) and by paper electrophoresis using polyacrylamide gel electrophoresis as a monitoring system. The amino acid composition of the purified peptide, determined after acid hydrolysis using an amino acid analyser (Dionex) is as follows: Asp, 2.08; Ser, 0.86; Glu, 2.02; Pro, 1.81; Gly, 2.01; Ala, 2.91; Met, 0.68; Ile, 0.95; Leu, 1.05; His, 1.04; Arg, 3.02. The peptide had an absorbance at 280 nm compatible with the presence of one tryptophan residue, giving a total composition of 20 amino acids. Aspartic acid was determined as the NH₂-terminal residue by identification of the corresponding dansyl derivative on polyamide sheets (A 1700; Schleicher & Schuell) after treatment with dansyl chloride. The primary structure of the icosapeptide was determined by manual Edman degradation and by degradation with carboxypeptidase Y; the sequence and details of the methods used are given in Fig. 2 legend.

Both chemical and immunological experiments were performed to determine the structural relationship between the isolated icosapeptide and the biosynthetically labelled material

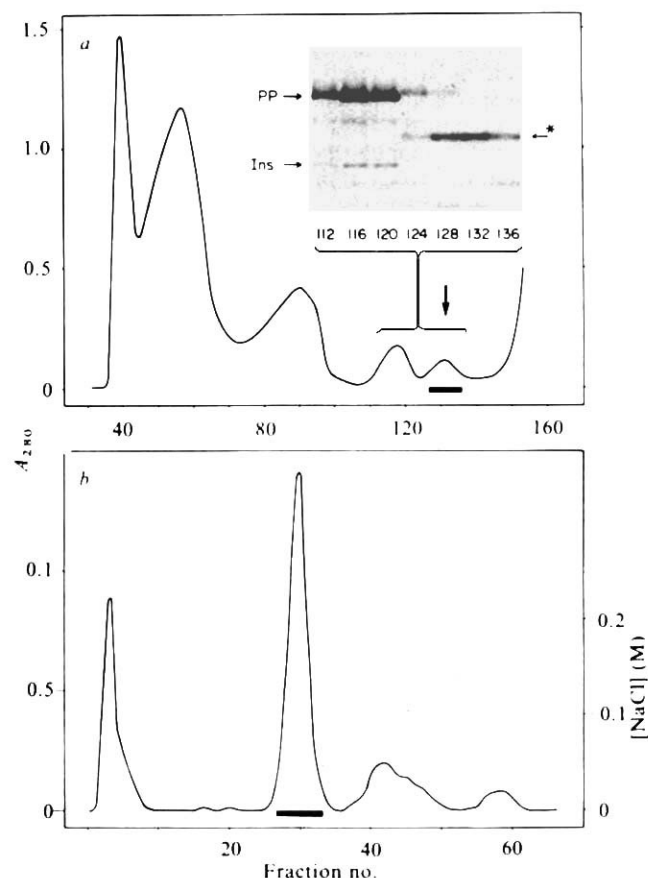


Fig. 1 Purification of the new peptide arising from pro-PP conversion. Duodenal lobes of pancreas (~20 g each) were excised from anaesthetized mongrel dogs and immediately homogenized in ice-cold acidified ethanol (0.1 M HCl in 68% ethanol, final concentrations). The homogenate was incubated overnight at 4 °C and centrifuged (40,000g, 25 min). The supernatant was neutralized and the peptides partially purified by ether precipitation²⁷. *a*, Gel filtration profile of proteins extracted from four pancreases on a 2.5 × 175 cm Bio-Gel P-30 column eluted with 3 M acetic acid. The vertical arrow corresponds to the elution position of the biosynthetically labelled fragment of the PP precursor (peptide III) obtained as described in Fig. 3 legend. The inset shows Amido black-stained peptides after polyacrylamide slab gel electrophoresis (10% acrylamide, pH 8.7; ref. 28) of aliquots from selected fractions of the area shown. The positions of PP and insulin (Ins) standards in the gels are indicated. The asterisk indicates the position of the peptide which was followed throughout the purification. The solid bar shows the fractions which were pooled and evaporated for further purification. *b*, Elution profile of gel-filtered peptides on DEAE-Sephadex. The ion-exchange column (1 × 4 cm) was equilibrated and eluted at 4 °C with 0.02 M Tris buffer pH 8.4 containing 6.5 M urea and with a NaCl gradient of 0–0.2 M (400 ml total volume). The solid bar indicates the fractions which were pooled and desalted over Bio-Gel P-2 using 3 M acetic acid. The peptide was further purified by paper electrophoresis on Whatman 1 or 3 paper in 30% formic acid and was eluted from the paper with 88% formic acid at 4 °C.

* To whom correspondence should be addressed at University Department of Clinical Chemistry, University of Copenhagen, Rigshospitalet, DK-2100, Denmark.

identified in PP cells. Endocrine cells were isolated from canine duodenal pancreas as previously described¹⁷ and purified on a Percoll gradient as described in Fig. 3 legend. The cells were incubated with ³⁵S-methionine or with ³⁵S-methionine plus ³H-leucine and the biosynthetically labelled peptides were studied initially by gel filtration. Three major radiolabelled peptides were detected (Fig. 3a). Peptide I is a biosynthetic precursor of PP which is converted during pulse-chase experiments into PP (peptide II) and its co-synthesized product (peptide III)¹⁷. The elution profile of labelled peptides from these freshly isolated, single cells was similar to that previously obtained from cultured, aggregated cells ('pseudoislets').

The biosynthetically labelled peptides identified in Fig. 3 were immunoprecipitated with anti-PP antibodies as well as antibodies raised against the icosapeptide (see Fig. 3 legend). The middle profile in Fig. 3a shows that the radiolabelled PP and its radiolabelled precursor are both precipitated by immune complexes prepared with anti-PP antiserum. The lower profile in Fig. 3a demonstrates that the radiolabelled PP precursor and the co-synthesized conversion product, peak III, are both absorbed specifically to immune complexes prepared using antibodies against the newly isolated icosapeptide. Thus, peptide I is specifically immunoprecipitated by antibodies against both PP and the icosapeptide, which shows that the precursor contains immunochemical determinants of both peptides. We have already demonstrated by peptide mapping that tryptic fragments of peptides II and III co-migrate with those of peptide I¹⁷. These two lines of evidence identify both the structure of the new icosapeptide and that of PP as occurring in the same peptide precursor.

The radiolabelled material in peak III (which was quantitatively immunoprecipitable by antibodies raised against the icosapeptide) separated into two peptides during polyacrylamide gel electrophoresis; one of these co-migrated with the icosapeptide whereas the other migrated more towards the anode at pH 8.7. To study the biosynthetic relationship between these peptides, the radiolabelled peptides obtained from peak III during a pulse-chase experiment were subsequently examined by polyacrylamide gel electrophoresis. As shown in Fig. 3b, upper profile, radioactivity appeared initially in the more anionic peptide, IIIa, during the pulse period. The middle and lower profiles in Fig. 3b demonstrate that during continued incubation in the presence of unlabelled amino acid, the radioactivity in the anionic peptide IIIa first increased (in

agreement with the processing of the precursor during the first 75 min of the pulse period¹⁷) and eventually decreased, while the radioactivity in the icosapeptide, peptide IIIb, increased. This shift in proportion of radioactivity between the two peptides, both of which are immunoprecipitable by antibodies against the icosapeptide, is consistent with the more anionic peptide being an intermediate in the biogenesis of the icosapeptide, the final product.

To determine the structural relationship between the two peptides eluting in peak III, they were purified by gel filtration and polyacrylamide gel electrophoresis and subjected to sequential Edman degradation. As shown in Fig. 2, ³⁵S-methionine was identified at position 5 in both peptides; in addition, ³H-leucine was identified at position 9 in the peptide which co-migrated with the icosapeptide. (Only peptide IIIb originated from a double label experiment.) The positions of these labelled amino acid residues align with the methionine and leucine residues in the NH₂-terminal sequence of the isolated icosapeptide. The identical position of the methionine residue in the NH₂-terminal sequence of the two radiolabelled peptides indicates that they differ at their COOH-terminus and that the intermediate form bears a short COOH-terminal extension to the icosapeptide sequence.

As the pancreatic polypeptide sequence has been localized at the NH₂-terminus of peptide I¹⁷, the results shown in Figs 2 and 3 indicate that the icosapeptide originates from the COOH-terminal part of the common precursor, from which the icosapeptide sequence is excised as an intermediate form (see Fig. 4); subsequent proteolytic events then refine the COOH-terminal region of the intermediate and result in the mature icosapeptide. The fact that the COOH-terminal residue of the icosapeptide is arginine suggests that an enzyme with a trypsin-like specificity is involved in cleavage of the intermediate form. Converting enzymes for peptide precursors usually require a dibasic amino acid sequence and often remove completely the converting-site residues. However, COOH-terminal arginyl residues are found in epidermal growth factor¹⁹ and nerve growth factor²⁰, both of which result from precursor cleavage by arginyl peptidases^{21,22}. In addition, the first two residues (Asp-Arg) are identical to the NH₂-terminal residues of angiotensin, which form the site for the specific conversion of angiotensins to des-Asp¹-angiotensin, angiotensin III²³. The icosapeptide derived from the processing of the PP precursor may be susceptible to a similar proteolytic modification.

Fig. 2 Amino acid sequence of the isolated icosapeptide and localization of methionine and leucine residues in biosynthetically labelled fragments of the PP precursor. Sequential manual Edman degradation, indicated by arrows facing right, was performed on 120 nmol of the icosapeptide by a method described elsewhere²⁰; all reagents were Sequanal grade (Pierce) and all organic solvents were distilled from ninhydrin before use. The phenylthiohydantoin (PTH) derivatives were identified on thin-layer silica sheets (Whatman LK5DF) developed with heptane/dichloroethane/propionic acid (50:20:13) and if necessary redeveloped in heptane/*n*-butanol/formic acid (75%) (50:30:9). PTH-arginine and PTH-histidine were identified by reaction with phenanthrenequinone³⁰ and Pauly's diazo reagent, respectively, both after spotting on Whatman 3 paper. The positions of leucine (9) and isoleucine (8) were confirmed by gas chromatography of the corresponding PTH derivatives. The COOH-terminal sequence was studied by degradation with carboxypeptidase-Y (Worthington) at 25 °C on 45 nmol of peptide with a molar enzyme/peptide ratio of 1:100 in 0.1 M *N*-ethylmorpholine brought to pH 7.0 with acetic acid. Norleucine (25 nmol) was added for standardization and aliquots removed after the following intervals: 0, 0.5, 1, 3, 5, 10 and 60 min, and 20 h. Free amino acid content was determined using an amino acid analyser. The arrows facing left indicate the amino acid residues which were placed in sequence from the time course of appearance during digestion with carboxypeptidase-Y; long arrows indicate that several amino acids were released too closely to be placed in sequence. No amino acyl phenylthiohydantoin derivative was identified after the 13th cycle of Edman degradation. However, the release of serine during carboxypeptidase digestion, as shown, is consistent with a serine at position 13. The histograms in the lower part of the figure illustrate the release of ³⁵S-methionine (closed columns) and ³H-leucine (open columns) during manual Edman degradation of the biosynthetically labelled PP precursor fragments (peptides IIIa and IIIb) obtained as described in Fig. 3 legend. These peptides were purified by polyacrylamide gel electrophoresis (15% acrylamide) at pH 8.7 (ref. 28) and by elution from the gel slices into 88% formic acid; the peptides were desalted over Bio-Gel P-2 using 3 M acetic acid. The icosapeptide (50 nmol) was used as carrier during both the purification and sequence analysis.

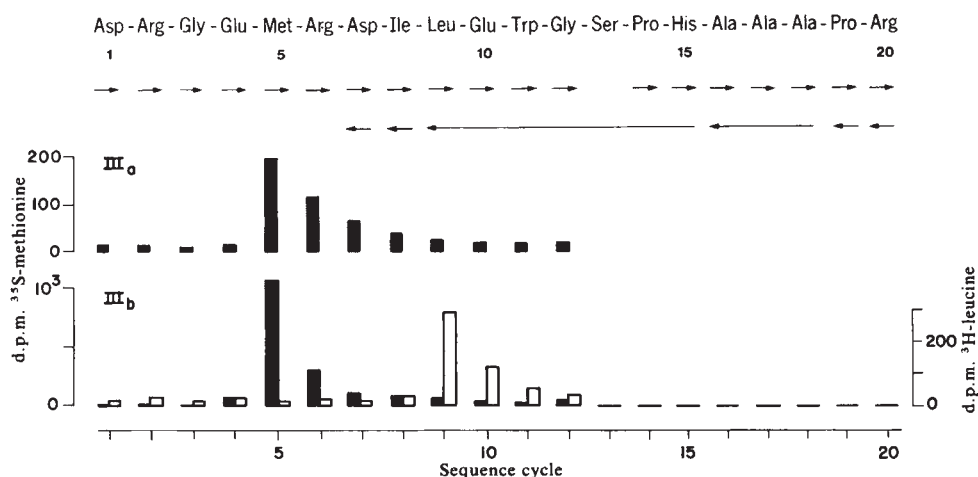
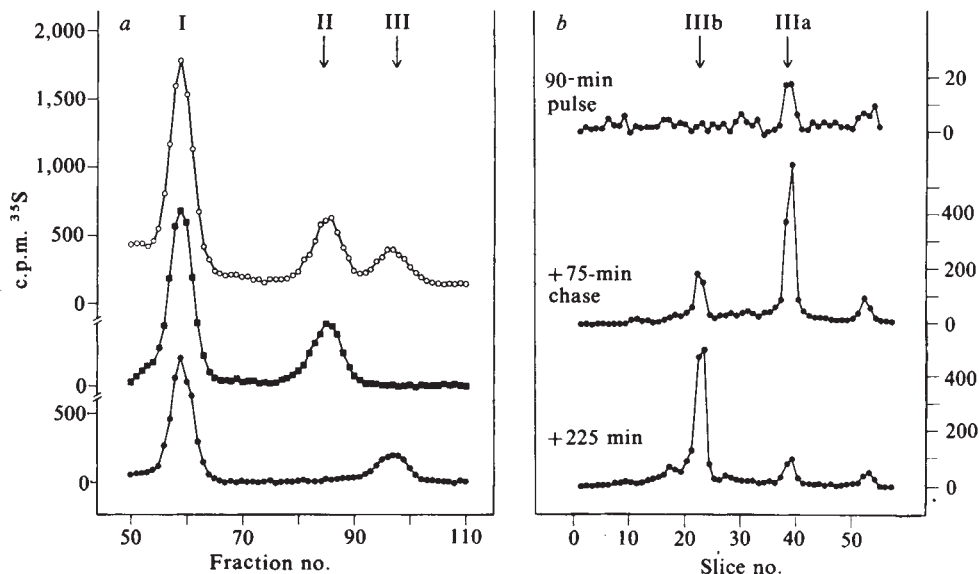


Fig. 3 Biosynthesis of PP and the icosapeptide in isolated endocrine pancreatic cells. Single cells were prepared from duodenal, canine pancreas by trypsin digestion in Ca^{2+} -free medium as previously described¹⁷. The endocrine cells were purified on a Percoll gradient and cells floating to the interface between layers having densities of 1.047 and 1.060 were used. The cells were incubated at 37°C in Eagle's basal medium (with Hanks' salts) lacking methionine, but containing 0.2 mCi ^{35}S -methionine (890 Ci nmol^{-1} ; Amersham). After 4 h, the cells were centrifuged and washed twice before extraction in 1 ml of 3 M acetic acid. The extract was then applied to a $1.5 \times 90 \text{ cm}$ Bio-Gel P-30 column and the column eluted with 3 M acetic acid containing 10 mg l^{-1} bovine serum albumin (BSA). **a**, Elution profiles of peptides appearing in the 12,000–1,000 M_r region. Aliquots from the column fractions were either counted directly or precipitated with preformed immune complexes prepared against PP or against the isolated icosapeptide as described previously¹⁷. (Anti-bovine PP antiserum (146–7) was given by Dr R. E. Chance, Lilly Research Laboratories, Indianapolis). Antibodies against the icosapeptide (isolated as indicated in Fig. 1) were raised in a rabbit (3202-2) injected subcutaneously with the peptide coupled through its α -amino group by difluorodinitrobenzene³¹ to BSA. Total radioactivity ($\circ$), radioactivity specifically immunoprecipitated with PP-antibody complexes ($\blacksquare$) and radioactivity specifically immunoprecipitated with icosapeptide-antibody complexes ($\bullet$) are shown. Specific immunoprecipitated radioactivity is defined as the difference in radioactivity precipitated in the absence and presence of excess immunocompetitor (bovine PP_{1–36} and highly purified icosapeptide, respectively). The void volume of the column corresponds to fraction 31 and PP has an elution position corresponding to peak II. **b**, Polyacrylamide gel electrophoresis profiles of labelled peptides obtained from peak III during a pulse-chase experiment. Endocrine cells were incubated with ^{35}S -methionine for 90 min (pulse period, top panel) followed by incubation with excess unlabelled methionine for 75 min (middle panel) or 225 min (bottom panel). The radiolabelled peptides were extracted and purified by gel filtration as described for **a** and peptides eluting in peak III were pooled and subjected to polyacrylamide gel electrophoresis in 8 cm-long gels containing 15% acrylamide. The gels were sectioned into 1.5 mm slices and peptides were eluted into 0.5-ml portions of 88% formic acid. The isolated icosapeptide co-migrated with peptide IIb in this system; the tracking dye bromophenol blue migrated to slice 53.



Although many large precursor molecules have been identified, only a few cases of the biogenesis of more than one stable peptide product from a single precursor have been documented. A classic example is the biosynthesis of insulin and C-peptide from proinsulin and the co-secretion of the two peptide products¹⁴. However, the C-peptide apparently serves only as a connector of the two chains of the insulin molecule. The secretion of equimolar amounts of glucagon and a fragment of the NH_2 -terminal part of the glucagon precursor has also been described¹⁶, but very little is known about the structure of this molecule and its possible function. Proopiomelanocortin, a precursor which contains the sequence of several bioactive peptides such as adrenocorticotrophic hormone (ACTH), melanocyte stimulating hormone (MSH), and β -endorphin^{15,24}, is a more complex example. Post-translational modification of proopiomelanocortin seems to differ between cell types and thus regulates the appearance of the different co-secreted peptides, for example, acetylation of α - NH_2 groups will activate α -MSH whilst inactivating the endorphins²⁵. Here we have reported the structure of the icosapeptide fragment of the PP precursor, which is stored in similar amounts to PP in the duodenal pancreas. Further studies will establish whether the storage and secretion of the icosapeptide or a related form can solve the

physiological enigma concerning the function of the PP-containing cells in the pancreas as well as in other tissues, for example the nervous system.²⁶

We thank D. Glock for help in obtaining the pancreatic tissue and P. Keim and R. Henrikson for performing the amino acid analysis. This study was supported by grants 512-10545 and 512-15837 from the Danish M.R.C. (to T.W.S.) and N.I.H. grants AM 18347 and AM 20595 (to H.S.T.). T.W.S. is a recipient of a postdoctoral fellowship from the University of Aarhus, Denmark, and H.S.T. is the recipient of Research Career Development Award AM 00145.

Received 10 August; accepted 19 October 1981.

- Kimmel, J. R., Hayden, L. J. & Pollock, H. G. *J. biol. Chem.* **250**, 9369–9376 (1975).
- Chance, R. E., Moon, N. E. & Johnson, M. G. in *Methods of Hormone Radioimmunoassays* (ed. Jaffe, B. M.) 657–672 (Academic, New York, 1979).
- Larsson, L.-I., Sundler, F. & Hakanson, R. *Diabetologia* **12**, 211–216 (1976).
- Malaisse-Lagae, F., Stefan, Y., Cox, J., Perrelet, A. & Orci, L. *Diabetologia* **17**, 361–365 (1979).
- Schwartz, T. W., et al. *Lancet* **i**, 1102–1105 (1976).
- Floyd, J. C., Fajans, S. S., Pek, S. & Chance, R. E. *Recent Prog. Horm. Res.* **33**, 519–570 (1977).
- Marco, J., Hedo, J. A. & Villanueva, M. L. *J. clin. Endocr. Metab.* **46**, 140–145 (1978).
- Lin, T.-M., Evans, D. C., Chance, R. E. & Spray, G. F. *Am. J. Physiol.* **232**, E311–E315 (1977).
- Taylor, I. L., Solomon, T. E., Walsh, J. H. & Grossman, M. I. *Gastroenterology* **76**, 524–528 (1979).
- Greenberg, G. R. et al. *Lancet* **ii**, 1280–1282 (1978).
- Bjornsson, O. G. et al. *Eur. J. clin. Invest.* **9**, 293–301 (1979).
- Adrian, T. E. et al. in *Gut Hormones* (ed. Bloom, S. R.) 265–267 (Churchill Livingstone, Edinburgh, 1978).
- Schwartz, S. S., Corkey, B., Williamson, J. R. & Rubenstein, A. H. *Endocrinology* **106**, 1178–1182 (1980).
- Rubenstein, A. H. et al. *Recent Prog. Horm. Res.* **33**, 435–468 (1977).
- Mains, R. E., Eipper, B. A. & Ling, N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3014–3018 (1977).
- Moody, A. J., Holst, J. J., Thim, L. & Linokaer Jensen, S. *Nature* **289**, 514–516 (1981).
- Schwartz, T. W., Gingerich, R. L. & Tager, H. S. *J. biol. Chem.* **255**, 11,494–11,498 (1980).
- Gersell, D. J., Gingerich, R. L. & Greider, M. H. *Diabetes* **28**, 11–15 (1979).
- Savage, C. R., Inagami, T. & Cohen, S. *J. biol. Chem.* **247**, 7612–7621 (1972).
- Angeletti, R. H. & Bradshaw, R. A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2417–2420 (1971).
- Berger, E. A. & Shooter, E. M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3647–3651 (1977).
- Frey, P., Forand, R., Maciag, T. & Shooter, E. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6294–6298 (1979).
- Peach, M. J. *J. Physiol. Rev.* **57**, 313–368 (1977).
- Nakanishi, S. et al. *Nature* **278**, 423–427 (1979).
- Hughes, J. *Nature* **278**, 394–395 (1979).
- Loren, I., Aluets, J., Hakanson, N. R. & Sundler, F. *Cell Tissue Res.* **200**, 179–186 (1979).
- Kenny, A. J. *J. clin. Endocr. Metab.* **15**, 1089–1105 (1955).
- Orstein, L. & Davis, J. B. *Disc Electrophoresis* (Distillation Products Industries, Rochester, New York, 1962).
- Peterson, J. E., Nehrlich, S., Oyer, P. E. & Steiner, D. F. *J. biol. Chem.* **247**, 4866–4871 (1972).
- Yamada, S. & Itano, H. A. *Biochim. biophys. Acta* **130**, 538–540 (1966).
- Tager, H. S. *Analyt. Biochem.* **71**, 367–375 (1976).

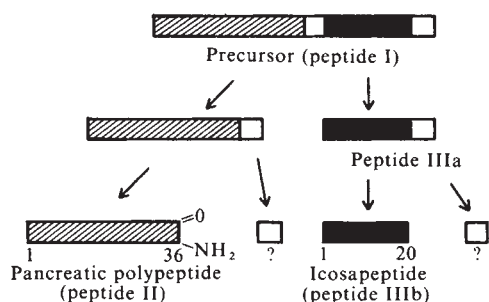


Fig. 4 Scheme for the organization of pancreatic polypeptide and icosapeptide within the common precursor and for the conversion of the precursor to its final stored products; PP and the icosapeptide are indicated by hatched and solid bars, respectively. The primary structure of PP has previously been localized to the NH_2 -terminus of the precursor by radiosequencing¹⁷. Roman numerals refer to the peptide designations used in Figs 2, 3.

Action of glutamate and aspartate analogues on rod horizontal and bipolar cells

R. A. Shiells, G. Falk & S. Naghshineh

Department of Biophysics, University College London, Gower Street, London WC1E 6BT, UK

Electrophysiological evidence indicates that transmitter release from retinal rod terminals occurs at a high rate in the dark and is reduced by light¹⁻⁶. The rod transmitter closes ionic channels (mainly sodium channels) in the subsynaptic membrane of a class of bipolar cells (on-centre bipolar cells) which depolarize in response to light^{4,7-9}. However, the transmitter opens sodium channels in the subsynaptic membrane of horizontal cells^{4-6,9,10}. We report here that several compounds, chemically related to the amino acids aspartate and glutamate, have similar post-synaptic actions to that of the rod neurotransmitter. Kainic acid¹¹, in micromolar concentration, hyperpolarizes rod on-centre bipolar cells, increasing their membrane resistance while depolarizing rod horizontal cells. Experiments with 2-amino-4-phosphonobutyric acid (APB), in which a phosphono group is substituted for a carboxyl group of glutamic acid, have further distinguished between different binding sites on the cells which make synaptic contact with rods. APB is an agonist at the on-centre bipolar cells, closing the same ionic channels as the rod neurotransmitter, but is without effect on rod horizontal cells. The evidence presented here suggest that the widely held view that glutamate or aspartate is the rod neurotransmitter needs to be re-examined.

Intracellular recordings were made in on-centre bipolar cells and horizontal cells of the virtually all-rod retina of the dogfish, *Scyliorhinus canicula*. Sectors of the dark-adapted eyecup were superfused with oxygenated elasmobranch Ringer's solution at pH 7.8 and ~16°C. Equilibration time, after switching perfusion fluids, was several minutes, limited by diffusion through the vitreous and retina. Test flashes of blue-green light, 15 ms in duration, were applied periodically. Full-field illumination was used. Methods of intracellular recording, identification of cells, resistance measurements and light calibration have been described previously⁹.

The effect of kainate on horizontal cells and on-centre bipolar cells is illustrated by typical records in Fig. 1a and b. Figure 1a shows a horizontal cell, depolarized by kainate at a concentration of 2 µM, the membrane potential increasing to +6 mV in the presence of 20 µM kainate. The upper trace (recorded simultaneously) shows the progressive decrease in the hyperpolarizing response to light of the horizontal cell with increasing concentration of kainate, the response to bright flashes being almost completely abolished by 20 µM kainate. On removal of kainate, the cell hyperpolarized to a final internal potential of -64 mV, close to its initial value in the dark (shown at the beginning of the lower trace); and, as seen in the upper trace, responses to light also recovered. The effects of kainate on horizontal cells are consistent with an action at the same sites as the rod neurotransmitter but do not rule out the possibility that the effect is exerted at a site presynaptic to the horizontal cell or at extrasynaptic sites on the horizontal cell. However, in other experiments synaptic activity was blocked by 2 mM Co²⁺, making the former possibility unlikely. The addition of Co²⁺ resulted in hyperpolarization of horizontal cells, to an internal potential between -100 and -120 mV, which was the same as the potential to which the cell could normally be driven by bright

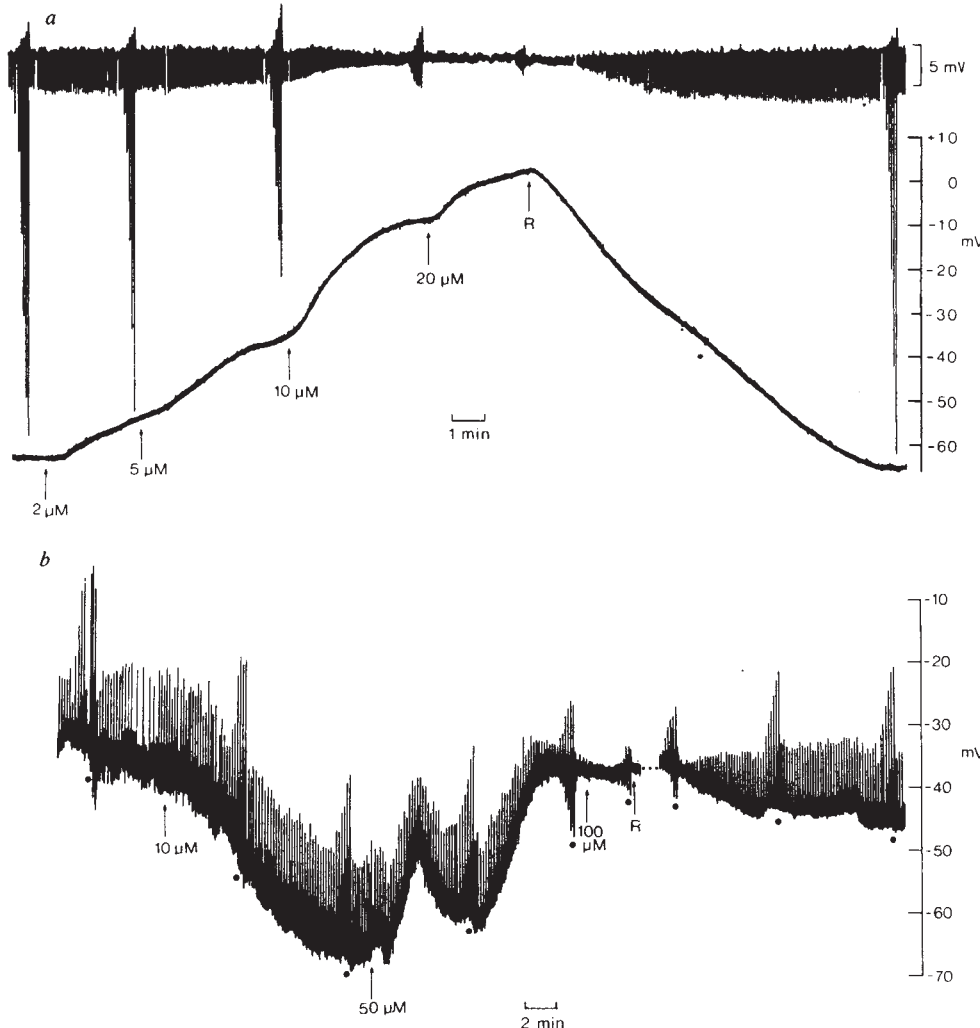


Fig. 1 The action of kainic acid on the membrane potential and response to light of a horizontal (a) and an on-centre bipolar cell (b). In this and succeeding figures, test flashes of constant low intensity were applied every 5–10 s; at intervals, the cell's light intensity-response relation (which appears as a staircase in response) was determined by increasing the light intensity by factors of ~2. a, The upper record shows the response of a rod horizontal cell to flashes, the response being represented as a displacement from the d.c. level in the dark. A sample-and-hold circuit was used to centre the d.c. level about an arbitrary zero before each flash⁹. The lower simultaneous record (at lower gain) shows the d.c. potential of the horizontal cell. Each of the constant test flashes bleached 0.87 rhodopsin molecules per rod (= 0.87 Rh^{**}). The brightest flash during each of the intensity-response determinations bleached 67 Rh^{**}. Perfusion with normal Ringer's solution is indicated by R. Normal elasmobranch Ringer's solution contained 280 mM NaCl, 3 mM KCl, 4 mM CaCl₂, 3 mM NaHCO₃, 0.5 mM MgCl₂, 350 mM urea, 10 mM glucose and 5 mM HEPES buffer pH 7.8. b Shows the effect of different concentrations of kainic acid on the membrane potential and response to light flashes of a bipolar cell. Test flashes, each bleaching 0.024 Rh^{**}, were applied throughout the record except when the light intensity-response relation was determined. Closed circles indicate the brightest flash of the series, bleaching 1.8 Rh^{**}. The dotted line indicates a break of 15 min in the record.

light, which presumably turns off transmitter release. The addition of 50–100 μM kainate resulted in depolarization to an internal potential of about zero, indicating that kainate acts directly on horizontal cells.

In contrast to its action on horizontal cells, low concentrations of kainate ($< 10 \mu\text{M}$) hyperpolarized on-centre bipolar cells. As shown in Fig. 1*b*, following the addition of 10 μM kainate to normal Ringer's solution, the bipolar cell hyperpolarized by 29 mV to an internal potential of -67 mV . The hyperpolarization was associated with an increase in input resistance from an initial value of 26 M Ω to 50 M Ω . After switching to higher concentrations of kainate, from 50 μM to 100 μM , complex transient effects on on-centre bipolar cells were observed, and responses to light flashes were significantly reduced. It is likely that complex effects on bipolar cells arise from a dual action of kainate, one a direct effect on the bipolar cell, the other arising from depolarization of horizontal cells which interact with bipolar cells. Figure 2 shows evidence that the interaction between horizontal and on-centre bipolar cells involves a synaptic mechanism. When synaptic transmission was blocked by Co^{2+} , 100 μM kainate monotonically hyperpolarized the on-centre bipolar cells, driving the membrane potential to values more negative than its normal value in the dark (Fig. 2*a*). Figure 2*b* shows the voltage displacements produced by applied current before, during Co^{2+} application and during application of Co^{2+} plus 100 μM kainate. Block of transmitter release by Co^{2+} depolarized the bipolar cell by 11 mV to a value of the internal potential which was slightly less than the maximum potential to which it could be driven by bright light. The input resistance decreased to 24 M Ω from its initial value of 39 M Ω . During the steady-state hyperpolarization by 100 μM kainate, the input resistance increased to 42 M Ω . The reversal potential for transmitter action was -15 mV , estimated from the intersection of the linear portion of current-voltage relations in normal Ringer's and during Co^{2+} block of transmitter release⁹. These results are consistent with the idea that the rod neurotransmitter closes ionic channels which are relatively selective for Na^+ . The reversal potential for kainate action, estimated from the intersection of the current-voltage relations during Co^{2+} block in the presence or absence of kainate, was estimated to be -13 mV . This result indicates that kainate also closes ionic channels which are relatively selective for Na^+ . The difference in values of reversal potential are within the error of the method.

The data may also be used to compare the changes in membrane potential with the resistance changes. If the channels closed by kainate and rod neurotransmitter had the same reversal potential, any change in resistance should be proportional to the change in membrane potential, but of opposite sign for a substance which closes channels⁹. This was found to be the case on the assumption that the resistance measurements were not more accurate than $\pm 5\%$. It was not possible in these experiments to distinguish whether the effect of kainate was at synaptic or extrasynaptic sites of on-centre bipolar cells. However, extrasynaptic open channels would constitute a shunt on the normal response to light and would reduce synaptic gain^{9,12}, whereas there is high gain in signal transmission between rods and on-centre bipolar cells⁹.

The putative transmitters, L-glutamate and L-aspartate, had similar effects on rod horizontal and on-centre bipolar cells to those of kainate, but were at least three orders of magnitude less potent, glutamate being about three times more active than aspartate. The required concentration of glutamate or aspartate was several orders of magnitude greater than expected for a neurotransmitter. As aspartate and glutamate may be transported into the cells and metabolized, the concentration at synaptic sites may be less than in the perfusate. Thus, concentration alone is not sufficient evidence against glutamate or aspartate being the rod neurotransmitter.

To examine further the possibility that glutamate or aspartate could be the rod neurotransmitter, substances which act as antagonists to these putative transmitters elsewhere in the nervous system were tested. One of these was APB, which binds

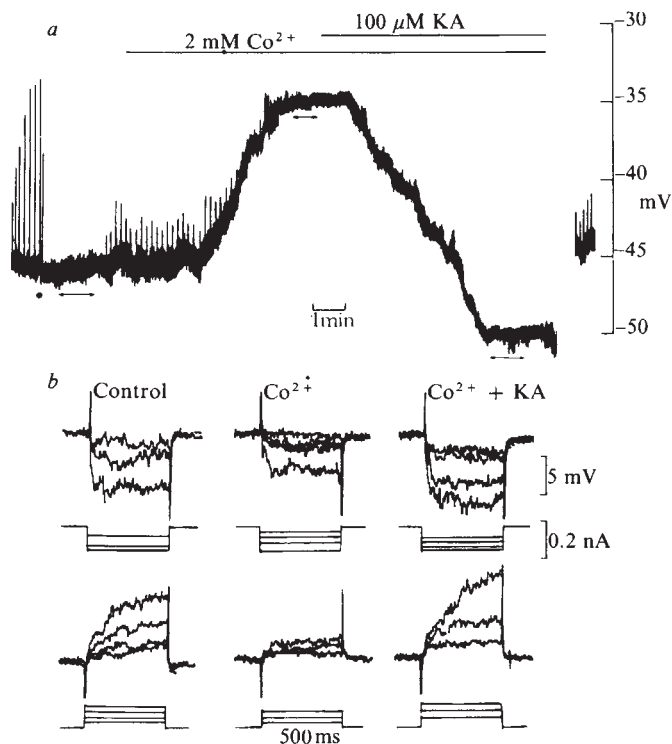


Fig. 2 Effect of kainic acid on membrane potential and input resistance of an on-centre bipolar cell when synaptic transmission has been blocked by Co^{2+} . *a*, Record of the membrane potential and response to 15-ms flashes given at 10-s intervals. At the beginning of the record, the control response to light flashes, increasing in intensity by factors of ~ 2 , is shown. The brightest bleached 4.4 Rh^{**}. The intervals indicated by arrows show the time when resistance measurements were made. The voltage displacements produced by the current pulses have been removed from the record for clarity. Low-intensity test flashes bleaching 0.053 Rh^{**} were applied throughout the rest of the recording. A 15-min break in the record following the removal of Co^{2+} and kainic acid is shown by a dotted line. *b*, Voltage displacements produced by hyperpolarizing and depolarizing current pulses of 500 ms duration, applied during the intervals shown in *a*. A single microelectrode was used for applying current and recording voltage. The voltage drop across the resistance in series with the cell's capacitance was balanced by a bridge circuit. The voltage displacement produced 100 ms after current onset, was used in constructing current-voltage curves. Because the records were noisy, the voltage displacements at 90–110 ms at a given applied current were averaged from six to eight records. The input resistance was estimated from the slope of the current-voltage relation in its linear region (displacements of $\pm 5 \text{ mV}$). Some of the apparent nonlinearity in the records is due to time-dependent changes in the resistance of the electrode during current flow. Cell input resistance in the absence of Co^{2+} and kainic acid (control) was 39 M Ω ; with 2 mM Co^{2+} , 24 M Ω ; with Co^{2+} plus 100 μM kainic acid, 42 M Ω .

competitively to glutamate-binding sites¹³ and acts as a competitive antagonist to glutamate^{13,14} at synapses where there is good evidence that glutamate is the excitatory transmitter^{15,16}. As shown in Fig. 3*a*, 50 μM APB reversibly hyperpolarized on-centre bipolar cells and blocked their responses to light, despite the increase in driving force for the light response; however, APB had no significant effect on horizontal cells (Fig. 3*b*). The experiment, illustrated in Fig. 4, in which synaptic transmission was blocked by Co^{2+} shows that the hyperpolarization of on-centre bipolar cells by APB results from its direct action on the bipolar cell. The increase in input resistance is evidence that APB acts by closing ionic channels. The reversal potential for transmitter action and APB action, estimated from the current-voltage relations, was the same, -31 mV for the cell illustrated. From these results we conclude that APB acts on the same sites of on-centre bipolar cells as those which are controlled by light. The selectivity of its action is due to different binding sites for the rod neurotransmitter on horizontal and on-centre bipolar cells.

APB is similar in size and charge distribution to glutamic acid and is ~ 500 times more effective than glutamate on on-centre bipolar cells. It is interesting that 2-amino-3-phosphonopropionic acid, which is shorter in length by one carbon

atom (and is the phosphono-analogue of aspartic acid), is also without effect on horizontal cells and is 50-fold less potent than APB on on-centre bipolar cells. The analogue, 2-amino-5-phosphonovaleric acid (APV), in which the chain length is increased by one carbon atom, had no effect on on-centre bipolar cells at a concentration of 0.1 mM (at least 20 times the concentration required for a detectable effect of APB). APV has been reported to be a potent antagonist of the excitatory effect of aspartate and glutamate in the mammalian and frog spinal cord¹⁷. The chemically related compound, D- α -amino adipate (D α AA) also acts as an antagonist to the excitatory actions of aspartate, and to some extent glutamate, on neurones of the central nervous system¹⁸⁻²⁰. We found that 5 mM D α AA had no effect on membrane potential or responses to light of on-centre bipolar cells. The racemic mixture (1 mM DL α AA) was also tested on rod horizontal cells and likewise was without effect. This is in contrast to the reported action of DL α AA on horizontal cells with an exclusive cone input, which provided some evidence for aspartate as a cone neurotransmitter²¹. Interestingly, the cone transmitter opens K^+ or Cl^- channels of on-centre bipolar cells in the retina of teleost fish whereas the rod transmitter closes Na^+ channels of the same bipolar cells⁸.

In view of the finding that glutamate and aspartate are only effective in considerably higher concentrations than required of APB and kainate, and that none of the antagonists tested here showed effects consistent with their expected actions, we suggest

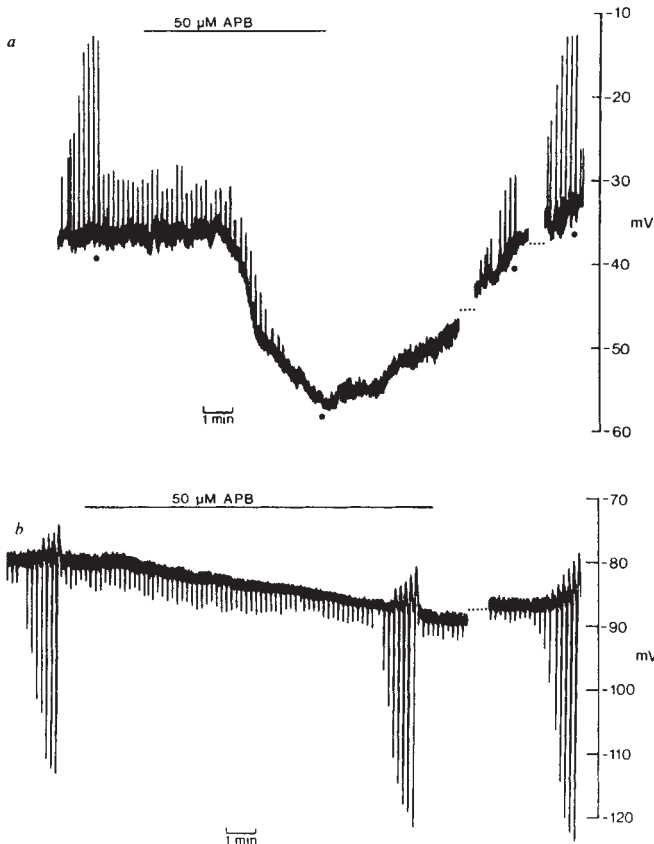


Fig. 3 Selective action of 2-amino 4-phosphonobutyric acid (APB) on on-centre bipolar cells compared with horizontal cells. *a*, The membrane potential and response to flashes of a bipolar cell. Closed circles indicate when the brightest of the test flashes was used in obtaining the relation between light intensity and response (4.4 Rh** except in the presence of APB when the brightest flash was 8 Rh**). The test flashes used throughout the rest of the record had an intensity of 0.024 Rh**. The first break in the record after washing out APB lasted 3 min, the second 11 min (dotted lines). *b*, The membrane potential and response to flashes of a horizontal cell. The small hyperpolarization seen after the addition of APB is probably due to a better seal of the electrode in the cell and not due to drug action, as it was not observed in other experiments. The brightest of the flashes in each of the intensity-response series was 136 Rh**. Each of the low-intensity test flashes bleached 0.47 Rh** except after the break in the record lasting 18 min, when the intensity was 0.2 Rh**.

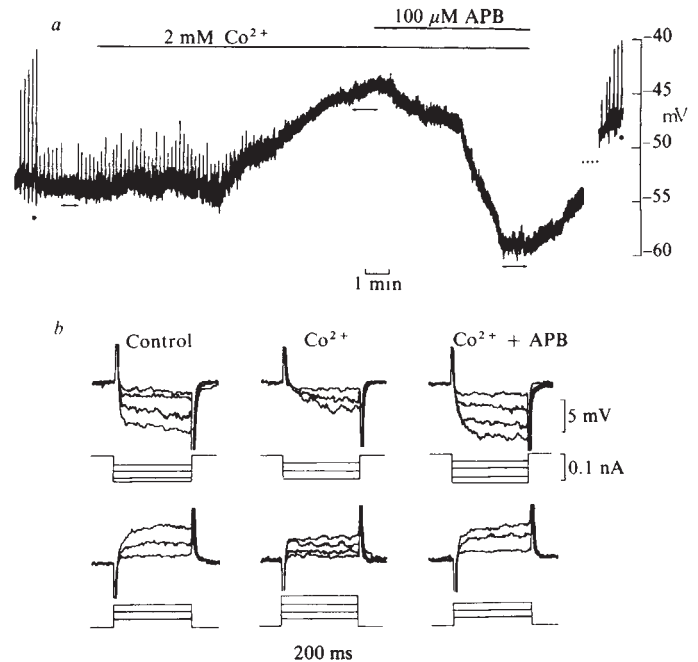


Fig. 4 Effect of APB on membrane potential and input resistance of an on-centre bipolar cell after block of synaptic transmission by Co^{2+} . *a*, Membrane potential and response to test flashes. The brightest flash at the beginning and end of the record (shown by closed circles) was of intensity 5.5 Rh**. Throughout the rest of the record, except during the intervals shown by arrows, the test flash was of intensity 0.029 Rh**. Resistance measurements were made during the arrowed intervals. The dotted line indicates a break of 12 min. *b*, Voltage displacements produced by 200-ms hyperpolarizing and depolarizing current pulses. Input resistance estimated in the way described in Fig. 2 legend: control, 42 MΩ; with 2 mM Co^{2+} , 25 MΩ; with Co^{2+} plus 100 μM APB, 54 MΩ.

that the candidacy of glutamate and aspartate as rod neurotransmitters requires re-examination.

The results raise the possibility that the rod transmitter acts at the same sites of rod horizontal and on-centre bipolar cells as those on which kainate acts. There is evidence that kainate may act at sites which are different from glutamate sites²²⁻²⁴. That the sites of transmitter binding on rod horizontal cells must be distinct from those of on-centre bipolar cells is shown by the selective action of APB. A similar selective effect of APB on on-centre bipolar cells of the mixed rod-cone retina of the mudpuppy has recently been described²⁵.

This work was supported by MRC and SRC grants. R.A.S. was in receipt of an MRC research studentship. Dr T. I. H. Brown participated in early experiments. We thank Drs J. Davies and J. C. Watkins for gifts of compounds.

Received 20 March; accepted 9 October 1981.

- Del Castillo, J. & Katz, B. *J. Physiol., Lond.* **124**, 586-604 (1954).
- Byzov, A. L. & Trifonov, J. A. *Vision Res.* **8**, 817-822 (1968).
- Toyoda, J., Hashimoto, H., Anno, H. & Tomita, T. *Vision Res.* **10**, 1093-1100 (1970).
- Kaneko, A. *Vision Res. Suppl.* **3**, 17-26 (1971).
- Falk, G. & Fatt, P. in *Handbook of Sensory Physiology* **7**, 1 (ed. Dartnall, H. J. A.) 200-244 (Springer, Heidelberg, 1972).
- Dowling, J. E. & Ripps, H. *Nature* **242**, 101-103 (1973).
- Toyoda, J. *Vision Res.* **13**, 283-294 (1973).
- Saito, T., Kondo, H. & Toyoda, J. *Vision Res.* **18**, 591-595 (1978).
- Ashmore, J. F. & Falk, G. *J. Physiol., Lond.* **300**, 115-150 (1980).
- Trifonov, Y. A., Byzov, A. L. & Chailahian, L. M. *Vision Res.* **14**, 229-241 (1974).
- McGeer, E. G., Olney, J. W. & McGeer, P. L. (eds) *Kainic Acid as a Tool in Neurobiology* (Raven, New York, 1978).
- Falk, G. & Fatt, P. *Vision Res.* **14**, 739-741 (1974).
- Cull-Candy, S. G., Donnellan, J. F., James, R. W. & Lunt, G. G. *Nature* **262**, 408-409 (1976).
- Dudel, J. *Pflügers Arch. ges. Physiol.* **369**, 7-16 (1977).
- Takeuchi, A., Onodera, K. & Kawagoe, R. *Proc. Jap. Acad. B* **56**, 246-249 (1980).
- Gerschenfeld, H. M. *Physiol. Rev.* **53**, 1-119 (1973).
- Davies, J., Francis, A. A., Jones, A. W. & Watkins, J. C. *Neurosci. Lett.* **21**, 77-81 (1981).
- Biscoe, T. J. *et al. Nature* **270**, 743-745 (1977).
- Hall, J. G., Hicks, T. P., McLennan, H., Richardson, T. L. & Wheal, H. V. *J. Physiol., Lond.* **286**, 29-39 (1979).
- Hicks, T. P., Guedes, R. C. A. & Creutzfeldt, O. D. *Brain Res.* **208**, 456-462 (1981).
- Wu, S. M. & Dowling, J. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5205-5209 (1978).
- Simon, J. R., Contrera, J. F. & Kuhar, M. J. *J. Neurochem.* **26**, 141-147 (1976).
- Walker, R. J. *Comp. Biochem. Physiol.* **55c**, 61-67 (1976).
- Hall, J. G., Hicks, T. P. & McLennan, H. *Neurosci. Lett.* **8**, 171-175 (1978).
- Slaughter, M. M. & Miller, R. F. *Science* **211**, 182-185 (1981).

BOOK REVIEWS

Biochemist extraordinary

Hans Kornberg

BIOCHEMISTRY has always been, and is likely to remain, an empirical science: a discipline in which knowledge accrues by experiment only. It is therefore not unexpected that major advances in the understanding of biological processes at the molecular level often follow the development of novel experimental techniques. Such developments occur when methods, currently available, are no longer adequate to provide answers to longer important questions; they usually involve the application to biological problems of technological advances made in other and often unrelated areas of science.

It is rare that one person has the combination of abilities that enables him to distinguish problems that are both important and capable of being resolved successfully, and to initiate and develop the novel means of studying them. It is even rarer to find this combination allied to so ruthless and single-minded a purpose that, for virtually the whole of his adult life, that person subordinates all other activities to the pursuit of his scientific goals. Such a person was Otto Warburg.

As Sir Hans Krebs reminds us in this masterly little book (there are only 81 pages of text, with a further 60 of documentation), Warburg's scientific achievements were immense. In the 1920s he adapted the manometric techniques, developed by Haldane and Barcroft for measurement of the amounts of gases in various body fluids, so that they became the main means not only for analysing a wide range of metabolites but also for measuring the rates at which cellular and sub-cellular events proceed. Warburg's investigation of the structure of the iron-containing component of the respiratory enzymes, that he had discovered by these techniques, led him to pioneer the use of photoelectric cells and monochromatic light sources to measure its action spectrum. This, in turn, led to the development of the UV spectrophotometer. Together with the Warburg manometer, the widespread use of this instrument permitted the pathways of cellular metabolism to be almost completely elucidated in the 1950s and 1960s. Incidentally, it was Warburg who discovered the identity of most of the coenzymes of respiration, including NAD, NADP, FMN and FAD; who discovered the change in light absorption at 340 nm consequent upon change in the oxidation state of NAD and NADP; and who isolated and purified numerous enzymes of carbohydrate metabolism including some — such as the first enzyme of the pentose

Otto Warburg: Cell Physiologist, Biochemist and Eccentric. By Hans Krebs. Pp.141. ISBN 0-19-858171-8. (Oxford University Press: 1981.) £10, \$24.95.



Otto Warburg at work, July 1966

phosphate pathway — not previously described. It is recorded that Warburg was judged worthy to receive a Nobel Prize on no fewer than three occasions over a period of nearly 20 years, for three separate scientific achievements.

But a price had to be paid for these successes. Throughout his astonishingly long and fruitful career, Warburg deliberately refused to "waste" time in pursuits that would not advance scientific discovery. Thus, apart from his interests in music and a wide knowledge of literature, Warburg shunned ties of family, of friendship or even of scientific fellowship; he did not teach, join committees or tolerate interruptions of his chosen routine. On one occasion, he refused an honorary DSc because he did not wish to devote time to collecting it. One would be tempted to see Warburg as an arrogant, self-righteous and monstrously egocentric despot in his laboratory, were it not for the thoughtful and sympathetic portrait that Sir Hans draws and that colours, with warmth and affection, this otherwise stark picture.

Sir Hans knew Warburg, as his student, colleague and (to some extent) confidant, for half a century. His accounts of the history and significance of Warburg's scientific work, and of the atmosphere in which it was carried out, make fascinating reading for any student of contemporary biology; his assessment of Warburg's

personality — his strengths, weaknesses, foibles, warts-and-all — makes equally absorbing reading for any student of human nature. One does not need to be an historian of science to feel a debt of gratitude to Sir Hans Krebs for producing what is as much an eminently readable work of scholarship as it is a labour of love.

Sir Hans Kornberg is Professor of Biochemistry at the University of Cambridge and a holder of the Otto Warburg Medal.

Creating confusion

John D. Barrow

The Creation. By P.W. Atkins. Pp.132. ISBN 0-7167-1350-0. (W.H. Freeman: 1981.) £5.95, \$9.95.

THE memorable Dr Pangloss was introduced to the public by his creator in order to lampoon a traditional but cyclopean perception of the natural world — the prejudice that there exists some benign and covert principle of optimality governing all things. The style of *The Creation* suggests that Pangloss must have had descendants who are not only alive and well but actively popularizing modern science. Ironically, they are no longer theists, but materialists advocating the pursuit of "extreme reductionism and militant rationalism".

The aim of this little book — and it is a good deal smaller than one might think since the text only appears on every other page with footnotes on the alternates — is to convince the lay reader that everything is as it is because it simply could not have been any different. Rather than take the traditional view that some Creator is necessary to explain the existence and arrangement of things, or (the equally traditional view) that he is not, the author argues that by extrapolating contemporary cosmological knowledge we can support a scenario wherein the Deity himself is confronted with what is now an increasingly familiar problem — redundancy. Roughly speaking, it is argued that nature behaves as the ultimate totalitarian state in which everything that is not forbidden is compulsory. By examining the laws and foundation properties of nature (as we now see them), Atkins tries to argue that different universes are not possible and that the traditional craftsman Creator — who carefully constructs universes by hand — is replaced by the computerized machinery of mathematical necessity.

This review was written before the death of Sir Hans Krebs last month. His *Reminiscences and Reflections*, to be published in January, will be reviewed in *Nature* at a later date.

Suspensions that we might be dealing with a new religion here are confirmed when on p. 7 we read that "the only faith we need for the journey is the belief that everything can be understood and ultimately there is nothing to explain". But, I suspect most readers' quarrel with the contents of Atkins's book would not be these aims or declarations of intent but the fact that the case he argues thereafter is also based upon faith or "knowledge" that usually does not yet exist. A Panglossian approach is followed whereby the author creates the type of future knowledge that would support his argument. This argument reminded me of mediaeval "proofs" for the existence of God, and suffers the same weakness that removed their ability to persuade. The disproof of any step in the arguments would not shake their authors' position one inch because the foundation of their belief really rested elsewhere.

That the bulk of the author's argument is speculation is not in itself a bad thing, but if we examine the way in which he handles known facts it ought to give us some guide as to the likelihood that this speculation is reliable. A few results of this rain-check were not too encouraging: he erroneously claims that it is the Lorentz signature of space-time that prevents time-travel; that the rate of biological evolution is exponentially sensitive to the fine structure constant (in fact, doubling the present value of the fine structure constant would, according to modern grand unified gauge theories, cause all protons to decay before stable stars form rather than increase the human evolution time to 10^{62} years!); and that the temperature of the environment is independent of the fine structure constant (it is actually determined by it).

To my mind, the author does not even get close to showing his case is a scientific possibility since his explanation for the inevitability of the creation of space-time *ex nihilo* requires an initial state of "almost nothing" — a dust of unstructured points (supplied by a well-known American relativist).

I am afraid that I would not recommend this book to the audience of laymen and students for whom it was created. Scientific popularizers and even scientific evangelists have a responsibility to wait until we know something about the creation of the Universe before telling the public everything about it. If you say, as the author does, that "complete knowledge is just within our grasp", to your professional colleagues it is quite harmless because they will simply laugh at you. To make such a statement in a popular book is totally misleading. This book will be far from enlightening and, I believe, will do little more than create confusion between the known, the unknown and the unknowable in the mind of the inquiring reader. □

John Barrow is a Lecturer in Astronomy at the University of Sussex.

The results of thinking small

Wolfgang K.H. Panofsky

The Nature of Matter. Edited by J.H. Mulvey. Pp.200. ISBN 0-19-8151151-5. (Clarendon/Oxford University Press: 1981.) £8.95, \$15.95.

GOOD popular expositions on particle physics are rare; this is one of them. It is a sequence of separate lectures given by prominent figures in the field at Wolfson College in 1980. In contrast to the common pattern in such series, the lecturers — at least occasionally — cross-refer to one another's talks, and there are other elements of coordination which make the volume rather better than most multi-author works.

Sir Denys Wilkinson opens the book with an excellent overview, starting with the development of atomistic concepts and ending with the efforts at grand unification of forces and particles. The subsequent discussions of particles and forces are similar to traditional popular presentations of the subject. I found the lecture by Llewellyn-Smith on manifest and hidden symmetries particularly useful, since it gives a comprehensible account of what is more generally called symmetry breaking where the usual layman's exposition leaves the reader in doubt why a broken symmetry is a symmetry at all.

The section by Abdus Salam, on unification of forces, is a straightforward, logical exposition beginning with a historical summary of how the many apparently disjointed forces in nature became understandable in terms of a smaller and smaller set. He then goes into more detail on the history of the unification of the gauge theories of electrodynamics and the weak interaction. His lecture is nicely interwoven with anecdotes; at the same time I fear the reader will gain relatively little understanding of the intellectual content of the theory — this, however, would be a difficult job in any circumstances. The lecture ends with an enumeration of the problems faced in the formulation of a grand unified theory and the alternative approaches to a quantum theory of gravitation.

John Ellis has written a fascinating chapter on the very large and the very small, which is mainly an outline of the extent to which our growing knowledge of the reactions of particle physics limits cosmological models.

I was somewhat less impressed by the experimental sections, in part because they are very brief and overshadowed by the theoretical lectures. The lecture by D.H. Perkins, "Inside the Proton", attempts to give an overview of the quark structure of nucleons and provides the usual quark and lepton tables. The fundamental neutrino scattering experiments which are of particular interest to the lecturer are elaborated, but the balance of the evidence

for hadron structure is covered only briefly. As a result the reader may get a misleading impression that the foundation on which the current theories of hadron structure rest is quite limited, while actually a firm experimental basis exists.

I have a similar reservation about the section on the tools of particle physics. The author, Sir John Adams, concentrates on the history and technical realization of proton circular accelerators. Since he is the creator of the CERN proton synchrotron and super proton synchrotron, this emphasis is fully justified. Yet, on reading this chapter, the layman will underestimate the richness of the arsenal of tools available to the experimentalist, in terms of accelerators, colliders and detectors. In particular — and here my own parochial view is evident — the essential role of electron-positron colliders, which have reaped a substantial fraction of recent discoveries, does not come through. This apart, the lecture includes some illuminating remarks on the social patterns that physicists have established in using the expensive tools of modern particle physics.

As a whole, *The Nature of Matter* does provide a readable account of various topics in elementary particle physics, ranging over the enormous breadth of dimensions to which modern theory applies. Yet what the layman will miss in the book is the appreciation that understanding of the forces and components of nature depends largely on experiment. The lecturers take little account of the impact of the major experimental discoveries — remember, for instance, the upset when the mu meson turned out *not* to be the carrier of the strong interaction predicted by Yukawa. Let us recall, too, the many unsuccessful attempts to explain theoretically the experimentally observed tau-theta paradox until Lee and Yang hit upon the right solution; the impact of the cosmic ray discovery of the "hooks and forks" which then became identified with the strange particles, from which then flowed the quark models of Gell-Mann and Zweig; the discovery of the Charmonium peaks which initiated the November Revolution of 1974; and finally the discovery of the third charged lepton which again was not predicted.

Somehow more has to be done in popular accounts of particle physics so that advances in the field are not conveyed as a sequence of events where theorists predict the future, machine builders construct suitable machines, experimenters design instruments and experimental arrangements to test the theory, and then theory is tested. That is not the way it is. □

Wolfgang K.H. Panofsky is Director of the Stanford Linear Accelerator Center, California.

Live issues: the classification of North American mammals

Malcolm C. McKenna

The Mammals of North America, 2nd Edn. By E. Raymond Hall. Two volumes, pp.1,181. ISBN Vol. 1 0-471-05443-7; ISBN Vol. 2 0-471-05444-5. (Wiley: 1981.) \$59.85, £33.25 per volume; two-volume set \$106.40, £59.15.

E. RAYMOND Hall and Keith R. Kelson's compilation of the species-level taxonomy and geographical distribution of living North American mammals, published in 1959, has long been out of print. Now Hall has produced a partly corrected second edition, updated to 1977 and even beyond, but the new version is still marred by many jarring anachronisms.

As before, the work is divided into two volumes. After a brief introduction, the first volume begins with marsupials and then marches through the insectivores, bats, primates, edentates, lagomorphs and the first half of a long section on rodents. The second volume completes the rodents and then moves to the cetaceans, carnivores, sirenians, perissodactyls and artiodactyls. In all of the entries, attention is focused on the traditionally-studied details of coat colour and measurements of the body, limbs, ears and tail — the sort of information one needs when setting out to match a stuffed mouse with specimens in the neatly arranged rows of museum trays. Many illustrations and occasional discussions of skull morphology are also provided, although, for me, details of some dentitions are depicted at too small a scale. There are helpful indexes, a glossary and a long bibliography, and a "how to do it" section deals with collection of specimens, preparation and data preservation.

I believe that the work is still flawed by its typological, philatelic approach — the preoccupation with segregating on the basis of a few selected criteria, mainly Rubicons of size, what might as well be inanimate, history-bereft objects. Although dichotomous keys are employed frequently, these have little phylogenetic rationale and do not deal with morphocline polarities as would be the case if the organisms were treated cladistically. No attempt is made to weed out shared-primitive characters. Worse, in the keys geographical distributions are all too often used in place of biological characters — just as stratigraphical position does not necessarily yield the correct identification

of fossils, geographical position does not necessarily give us the correct taxonomic allocation of living organisms. Biological taxa should be recognized by biological criteria alone. Let the stratigraphical and biogeographical chips fall where they may.

According to Hall, about 400 species and subspecies have been added to the living North American mammalian fauna since 1959, bringing the present total up to 3,607. However, although myriads of named (and presumably approved) subspecies are listed and referenced, morphological evidence for their recognition is not presented. Rather, maps depict arbitrary boundaries between localities that produce "marginal specimens" of the putative subspecies. The characterizing features of these — if genuinely biological, and not merely geographically arbitrary or a relic of overly split typological taxonomy — must be sought either in the original literature or elsewhere. Perhaps, in some long-distant future revision, the subject of intraspecific variation can be handled quantitatively. Hall has made a start, but the job is immense.

Finally, Hall's treatment of higher taxonomic categories should be taken with a grain of salt. This is because he deals with members of the living fauna as though they had no history: he does not differentiate between primitive features and phylogenetically significant shared-derived

ones. For example, he seems to be unaware that characters such as an inflected mandibular angle or epipubic bones are not merely metatherian hallmarks but characterize early eutherians as well. Nor does he record that jugal bones reach the glenoid fossa in paenungulates, not just in the Metatheria. Interested readers should also run through the list of mainly primitive features supposedly characterizing Polyprotodontia, Didelphidae or Insectivora (the latter stretched to include the conceptually indigestible macroscelidids). Hall's concept of the "Order Pinnipedia" and his separation of living sloths from megalonychid ground sloths by interlarding the Myrmecophagoidea into their midst will also raise palaeontological and cladistic eyebrows alike.

Hall's work will continue to be useful in spite of its conceptual shortcomings and his apparently stubborn refusal to modify adequately some sections of the first edition which many, even most, mammalogists found objectionable — those dealing with bears, *Dama* and *Lasiurus*, for example.

After all these grumblings, do I recommend the monograph? Yes, of course. After all, what better summary do we have? □

Malcolm C. McKenna is in the Department of Vertebrate Paleontology at the American Museum of Natural History, New York.

An abundance of beetles

M.E.G. Evans

The Biology of the Coleoptera. By R.A. Crowson. Pp.802. ISBN 0-12-196050-1. (Academic: 1981.) £58, \$139.50.

WE LIVE in the age of the obscure beetle. Flowers, fleas, birds and butterflies are all much more obvious organisms, but there are more species in the order Coleoptera than there are in the class Angiospermae, and even, perhaps, than in the whole plant kingdom. Since this enormous species diversity is based upon a common body plan, the variations on this theme should be of great interest to evolutionary biologists as well as specialist entomologists. So where are the hosts of coleopteran biologists?

One reason why biological studies of beetles lag far behind their taxonomy may be a simple failure to appreciate the scope of the subject. What has been lacking until now has been a general biological text which deals with the numerous but scattered studies of beetles. It is just this sort of book which Dr Crowson has written; it is based on well over 1,000 references, including many from non-

English language literature, most of which is recent. However, it is more than a mere compendium of references. In most cases, the information on a topic has been summarized, digested and assimilated into a more general subject area in each of the 20 chapters of text, ranging from morphology and genetics to ecology and evolution. The basic approach is both functional and phylogenetic, although an extended treatment of classification is deliberately omitted. Nevertheless, a family classification is appended, and taxonomic groups in the text are given code numbers which refer to the phylogenetic chart following Chapter 20.

Inevitably, there are some blemishes. Over 300 figures are included, most of which are simple, clear, line drawings, but some have been drastically reduced for printing whilst a few are not reduced enough. Not surprisingly, some topics are discussed in less depth than others, but in most cases these differences in treatment reflect our different levels of comprehension of the subject. Indeed, this has enabled the author to point to obvious gaps

The Biochemistry of the Nucleic Acids, originally written by J.N. Davidson and first published in 1950, has recently appeared in a ninth edition. The new edition has been compiled by R.L.P. Adams together with four other authors at Glasgow University, and is a substantial revision of the last, 1976, edition. The book is published by Chapman & Hall and costs £8.50 in paperback.

in our knowledge of the Coleoptera over the whole range of their biology, making one of the most valuable features of the book.

Taken together with its encyclopaedic coverage, these indications of our ignorance will enable *The Biology of the Coleoptera* to provide many potential researchers with a starting point, a basic literature survey and a summary of current knowledge. This is one of the most important books on beetles to be published this century, and it is the first modern text on their biology. I hope it will be bought by many entomologists, evolutionary biologists and science libraries. □

M.E.G. Evans is Senior Lecturer in Zoology at the University of Manchester.

Puzzles of perception

Oliver Braddick

Direct Perception. By Claire F. Michaels and Claudia Carello. Pp.200. ISBN 0-13-214791-2. (Prentice-Hall: 1981.) £13.50, \$24.30.

THE perceptual psychologist J.J. Gibson (1904–1979) wrote a series of books which crystallized a theoretical stance he called "ecological optics" and "direct perception". The books were widely read and many details from them became absorbed as standard textbook material, but it seemed for many years that Gibson's basic position was a lonely, almost ignored one, outside the mainstream of thinking on perception. It is still outside the mainstream, but in the years just before and since Gibson's death it has gathered such a vigorous, even strident, group of younger advocates that no one in the field has an excuse for being unaware of it. There has been a particularly zealous Gibsonian cell in and around the University of Connecticut, two of whose members have produced this short, clear, didactic book.

The view against which the authors argue is that the perceiving organism takes in at the sense organs a collection of "raw data" whose relation to objects in the external world is distant and often incomplete, and that the task of perceptual psychology is to discover the computations by which information about the external world is obtained or inferred from these data. The alternative they propose is that perception has to be considered as a function of a system consisting of organism and environment, that the ecologically relevant properties of the environment are present as higher-order invariants in the pattern of energy reaching the senses, and may be detected because the perceptual system "resonates" directly to these higher-order invariants, rather than

computing or inferring anything. Some other advocates of this position have presented it in a manner that appeared to relish shocking, mystifying or finessing more conventional views; here the argument is more orderly and well paced, with its progress charted in summaries of each chapter and at the end of the book. An appendix, "Discussion and Debate", which deals with some queries and criticisms of the argument in question-and-answer form, is very helpful in clarifying just what the authors' position is, although as a debate it is distinctly stage-managed.

It is a valid criticism of much work in sensory and perceptual psychology that it does not give enough thought to what perception is *for*, or to whether the stimulus variables that are manipulated in perceptual experiments bear any relation to the variables the perceiver needs to use. The opponent that the book sets up is therefore not a straw man, but he lacks flesh in a way that the Gibsonians blithely ignore. The information-processing tradition is centrally interested in problems of mechanism; not only (indeed, not enough) in what job perceptual systems do, but also in how they work. For instance, Michaels and Carello argue that the direct perception approach undercuts the two "puzzles in binocular vision" which have been "an object of discussion for three millennia" — how can a pair of two-dimensional retinal images yield a 3-D percept, and how can it be single? No awareness is indicated that the major binocular puzzle that has interested scientists in the last two decades is not of this kind but is "What kind of mechanism finds pairs of corresponding elements in the two eyes' images?". Now analysis of ecological optics might suggest that this question should be posed in a somewhat different form, but identification of the relevant invariants in the optic array does not abolish the question, how does the system work which "resonates" to them? It is difficult to see how such an answer can be cast except in terms of signals relayed by the sense organs and operations performed upon these signals.

For this reason, scientists working in the information-processing and psychophysical traditions are likely to be impatient with much of *Direct Perception*. What they would find it hard to deny is that most of the experimental work generated by the direct perception school has been interesting — something they would have to be very partisan to assert of their own tradition. It is perhaps a pity that Michaels and Carello did not make their book 50 pages longer and give us a more detailed awareness of how their approach bears scientific fruit — in the experimental and analytical work of Lee, Shaw and Turvey, among others. □

Oliver Braddick is a Lecturer in Experimental Psychology at the University of Cambridge.

Bones into stones

Mark Newcomer

Life History of a Fossil: An Introduction to Taphonomy and Paleoecology. By Pat Shipman. Pp.222. ISBN 0-674-53085-3. (Harvard University Press: 1981.) \$25, £14.

PALAEONTOLOGISTS are no longer content to classify, weigh, measure and tabulate fossil bones, and are bringing their subject out of the museum through studies of the processes by which the fossil record forms (taphonomy) and through reconstructions of extinct animal communities in their environmental setting (palaeoecology). Recent work in these fields, which to some degree parallels that of archaeologists who attempt to see their sites and artifacts in their geographical context, is the subject of Pat Shipman's lively text.

Much of the book concentrates on the principles, methods of study and results of taphonomic studies of (mainly) African vertebrates, reflecting the author's research interests and competence. She treats her material critically, steering a cautious course through such murky waters as the spiral fracture debates, and stresses the value of suitable quantitative methods for many taphonomic and palaeoecological problems; her comments on excavation procedures and "living floors", however, seem a little naive, especially in light of all the recent publications on these matters. (Why does the long reference list contain only English-language publications? Curious in a book intended as a text.) Without sounding arrogant, I think it is fair to say that standards of excavation have generally been higher in archaeology than in palaeontology, with the fossil hominid hunters perhaps the most culpable, relying on the spectacular nature of their finds to obscure excavation methods which are often mediocre or worse. Similarly, the author could have gone more deeply into the role of hominids in creating both the types and spatial patterns of refuse on sites, given recent experimental work and ethnographic studies by Binford and Yellen, among others.

Intended for "advanced undergraduate courses in anthropology, palaeontology, and evolutionary biology", as well as professionals in these fields, *Life History of a Fossil* will serve as an excellent introduction to taphonomy and palaeoecology. It should also reach a much wider audience with less specialized interests in vertebrate evolution, to whom the lessons to be learned from humble, often broken, bones will come as a pleasant surprise. □

Mark Newcomer is a Lecturer in Palaeolithic Archaeology at the Institute of Archaeology, London.